Why is Africa losing ground against malaria? p. 1139

The various impacts of COVID-19 on children p. 1144

Global warming hastens climate-tipping transitions p. 1171

Science

$15
9 SEPTEMBER 2022
science.org

AAAS

DOMESTICATING DONKEYS

Donkeys emerged in Africa more than 6000 years ago p. 1172

BII | Prize for Science | Innovation

Bringing research to life–and science to market

Behind every life-changing solution is an entrepreneurial scientist–a creative mind who proved an idea in the lab and dared to carry it out in the world.

To encourage more scientists to translate their research, BioInnovation Institute (BII) and *Science* present a new annual award.

Our three winners will have their essays published in *Science* magazine and will be invited into BII's entrepreneurial ecosystem. In addition, the Grand Prize winner will receive a prize of USD 25,000, and each runner-up will receive USD 10,000 at a grand award ceremony in Copenhagen, Denmark.

The call for applications has just opened. Apply before November 1, 2022.
www.bii.dk/scienceprize

See film from the grand award ceremony 2022

Benedetto Marelli, Grand Prize Winner 2022

Presented by BII & *Science*

BII BioInnovation Institute Science

Apply before November 1, 2022
www.bii.dk/scienceprize

CONTENTS

9 SEPTEMBER 2022 • VOLUME 377 • ISSUE 6611

A researcher examines a mosquito trap during a malaria study in Mozambique.

PHOTO: CYNTHIA R. MATONHODZE

NEWS

IN BRIEF

1130 News at a glance

IN DEPTH

1132 Twisty device explores alternative path to fusion
Revamped German stellarator should run longer, hotter and compete with tokamaks *By D. Clery*

1133 New tech law offers billions for research
CHIPS act will fund microelectronics innovation and training through large partnerships *By J. Mervis*

1135 Warming of 1.5°C carries risk of crossing climate tipping points
Scientists call for concerted effort to forecast points of no return for ice, weather patterns, and ecosystems *By C. O'Grady*
RESEARCH ARTICLE P. 1171

1136 California EV rules jolt battery science
Move to phase out gas-powered cars will force progress toward faster charging batteries *By R. F. Service*

1137 Fauci looks back—and ahead
Loved and hated, NIAID's chief plots life after government *By J. Cohen*

FEATURES

1139 Lingering fever
The battle against malaria in Africa has stalled. Can research in Mozambique explain why—and how to get it back on track?
By L. Roberts
PODCAST

INSIGHTS

PERSPECTIVES

1144 Viewpoints: COVID-19
COVID-19 and children

1150 Understanding myalgic encephalomyelitis
Myalgic encephalomyelitis and Long Covid have overlappping presentation
By S. M. Gradisnik and N. Eaton-Fitch

1152 Chiral light sources get a helping hand
Resonant metasurfaces pave the way for more compact sources of pure chiral light *By A. Forbes*
REPORT p. 1215

1153 El Niño's role in changing fauna
Past faunas show how El Niño frequency affects eastern Pacific community structure
By D. H. Sandweiss and K. A. Maasch
REPORT p. 1202

1155 Scaling brain neurogenesis across evolution
A genetic change could explain increased cortical neurogenesis in modern humans *By B. Malgrange and L. Nguyen*
RESEARCH ARTICLE p. 1170

1156 Three types of planets orbit red dwarfs
Precise densities of red dwarf exoplanets help distinguish potential "water worlds"
By J. Teske
REPORT p. 1211

POLICY FORUM

1158 Exclusion cycles: Reinforcing disparities in medicine
Clinical practice, data collection, and medical AI constitute self-reinforcing and interacting cycles of exclusion *By A. Bracic* et al.

BOOKS ET AL.

1161 Connecting for success
Interventions that increase our sense of belonging can help us confront new challenges *By M. Lieberman*

CONTENTS

1156 & 1211

1162 A tale of troubled waters
Vietnamese refugees faced misdirected rage for a marine ecosystem ravaged by industry, argues a journalist *By M. E. Hannibal*

LETTERS

1163 Joint protection of a crucial reef system
By Y. Sun et al.

1163 Adaptive wildfire mitigation approaches
By D. Lindermayer et al.

1164 Restore China's coastline from the ground up
By X. Wang et al.

RESEARCH

IN BRIEF

1165 From *Science* and other journals

REVIEW

1168 Land use change
Disentangling the numbers behind agriculture-driven tropical deforestation
F. Pendrill et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABM9267

RESEARCH ARTICLES

1169 Molecular biology
Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT *H. Ehara* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABP9466

1170 Neuroevolution
Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals *A. Pinson* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:DOI.ORG/10.1126/SCIENCE.ABL6422
PERSPECTIVE P. 1155

1171 Climate change
Exceeding 1.5°C global warming could trigger multiple climate tipping points
D. I. Armstrong McKay et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:DOI.ORG/10.1126/SCIENCE.ABN7950
NEWS STORY P. 1135

1172 Domestication
The genomic history and global expansion of domestic donkeys *E. T. Todd* et al.

1180 Cancer
Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling
J. M. Chan et al.

1192 Correlated materials
Systematic electronic structure in the cuprate parent state from quantum many-body simulations *Z.-H. Cui* et al.

REPORTS

1198 Mesoscopic physics
Isolated ballistic non-abelian interface channel *B. Dutta* et al.

1202 El Niño
El Niño frequency threshold controls coastal biotic communities
J. M. Broughton et al.
PERSPECTIVE P. 1153

1206 Evolutionary biology
Evolutionary gain and loss of a pathological immune response to parasitism
J. N. Weber et al.

1211 Exoplanets
Density, not radius, separates rocky and water-rich small planets orbiting M dwarf stars
R. Luque and E. Pallé
PERSPECTIVE P. 1156

1215 Metasurfaces
Chiral emission from resonant metasurfaces
X. Zhang et al.
PERSPECTIVE p. 1152

1218 Solid-state physics
Spin-selective tunneling from nanowires of the candidate topological Kondo insulator SmB_6
A. Aishwarya et al.

1223 Hydroformylation
Hydroformylation catalyzed by unmodified cobalt carbonyl under mild conditions
B. Zhang et al.

1227 Solar cells
Transporting holes stably under iodide invasion in efficient perovskite solar cells
T. Wang et al.

DEPARTMENTS

1129 Editorial
Protect the vulnerable from monkeypox
By B. Titanji

1234 Working Life
A walk in the office
By N. Eisenhauer

ON THE COVER

A donkey from Petra, Jordan, carries goods through the rugged landscape, as millions of others have done for millennia. Genetic analysis tracks their journey from a single domestic origin in Africa as early as 7000 years ago to rapid expansions through Eurasia 2500 years later. Ancient genomes also reveal trans-Mediterranean exchange, giant Roman beasts, and previously unknown diversity in the Levant. See page 1172. *Photo: Brit Büchner/EyeEm via Getty Images*

Science Careers 1233

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Copyright © 2022 by the American Association for the Advancement of Science.

The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

ILLUSTRATION: ESO/M. KORNMESSER/CC BY 4.0

EDITORIAL

Protect the vulnerable from monkeypox

Recent declines in monkeypox infections have elicited comments such as "cautiously optimistic" and "we're turning a corner." Indeed, this trend in large cities in the United States and Europe, as well as a global decline in new cases of 21% (according to the World Health Organization), is encouraging. However, if we don't prioritize and protect the most vulnerable populations, the outbreak will not end any time soon.

Since May 2022, monkeypox has rapidly spread to 102 countries, with over 50,000 cases globally. Despite gaining international attention recently, the virus has ravaged West and Central Africa for many decades. Warning signs were ignored and as a result, the world knows little about the pathogen's biology and the potential of vaccines and antiviral treatments to contain its spread.

Cases of monkeypox show a striking parallel with HIV. Widespread, rapid transmission of both viruses first occurred in sexual networks of gay, bisexual, and other men who have sex with men (GBMSM). How should we balance a pressing research agenda for monkeypox while safeguarding the populations most at risk?

Engaging vulnerable communities early is an important lesson from the HIV epidemic. In Atlanta, Georgia, where I provide care to patients with monkeypox, 67% of cases have been persons living with HIV, and a notable 71% are Black GBMSM. Only 46% of vaccine doses administered in Georgia have gone to Black GBMSM, who are clearly the most highly affected group. The fear of reliving the stigma and uncertainties experienced in the early days of the HIV epidemic is palpable in almost every monkeypox patient I've encountered. Justified concerns about equitable access to treatments and vaccines among minority groups compound these fears. Also evident is the fierce desire of individuals to protect their community. Empowering them to do so requires allowing community members to lead conversations on what should be prioritized and how best to engage with citizens to ensure their needs are met.

As with other infectious diseases like HIV and COVID-19, a one-size-fits-all approach will likely not be effective. Interventions may affect the spread of monkeypox in GBMSM networks differently than in other communities. In West and Central Africa, virus transmission (following an animal-to-human zoonotic jump) has mainly been among young children and women who serve as primary caregivers. Thus, tailoring communication to the needs of specific communities is essential. For example, there has been public health hesitation in highlighting the important role of transmission through sexual contact in ongoing outbreaks. Acknowledging the role of sex in the spread of monkeypox among GBMSM is not stigmatizing but rather allows individuals to better evaluate their own risk and ways to help mitigate it.

During the early years of the HIV epidemic, stigma and discrimination against GBMSM stifled research in targeted prevention. Failure of policies to address HIV were blamed on this community, further ostracizing them. The same is happening with monkeypox today. The lesson learned from HIV is that legislation, policy, and funding are needed that protect vulnerable communities from stigma. Laws that criminalize homosexuality in some countries in West and Central Africa, for example, will make it challenging to conduct studies aimed at understanding monkeypox spread in GBMSM social and sexual networks.

> "What about equitable access to the fruits of monkeypox research?"

The need to safeguard vulnerable groups must be balanced with the risk of overprotecting them. For example, the exclusion of young children and pregnant persons from vaccine and drug studies for their "protection" often results in insufficient data on interventions that could be beneficial to them. What about equitable access to the fruits of monkeypox research? In the past 3 months, the rollout of vaccines and investigational treatments for monkeypox almost exclusively occurred in North America and Europe. Yet much of what is known about the virus is informed by research in Africa on African populations. Participation in research has not guaranteed access to interventions in a time of crisis. This has to stop.

The health disparities seen during the spread of HIV prevailed during the COVID-19 pandemic and are again apparent with monkeypox. Unless vulnerable populations become an integral part of tackling monkeypox globally, from research participation to accessing interventions, the world will likely make the same mistakes again.

–Boghuma Titanji

Boghuma Titanji is an assistant professor in the Division of Infectious Diseases, Emory University School of Medicine, Atlanta, GA, USA. btitanj@emory.edu

PHOTO: COURTESY OF BOGHUMA TITANJI

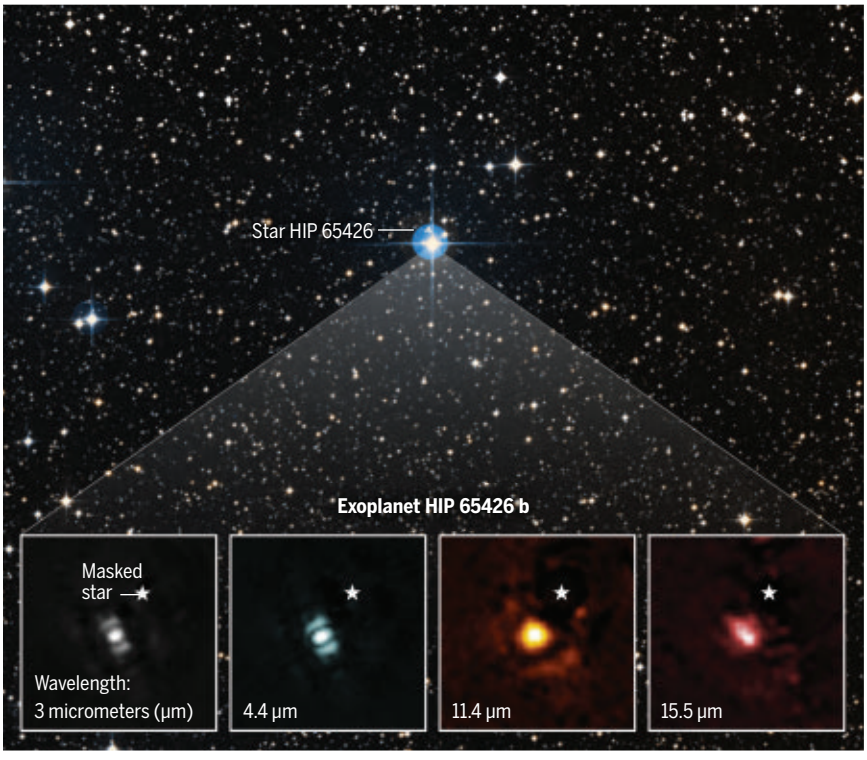
10.1126/science.ade7115

NEWS

> “No one goes to jail.”
>
> **Luciano Evaristo,** former chief inspection officer of Brazil’s environmental law enforcement agency, in *The Washington Post*, about people charged with deforesting the Amazon. Observers also decry a lack of inspectors and indifference by the national government.

IN BRIEF

Edited by **Jeffrey Brainard**



An exoplanet was imaged at various wavelengths (insets); the white pentagram marks the parent star.

ASTRONOMY

Webb telescope snaps its first exoplanet

The James Webb Space Telescope has taken its first images of a planet beyond our Solar System, opening a new chapter in the discovery and study of exoplanets. The young gas giant, seven times the mass of Jupiter, is captured still glowing hot after its formation, its infrared radiation having traveled 350 light-years to reach Webb’s gold-plated mirrors. Webb relied on an optical mask to block the overwhelming glare of the star and image the planet, HIP 65426 b, at multiple infrared wavelengths, astronomers reported in a preprint. About 20 other exoplanets have been imaged directly, though telescopes on Earth and in space. But Earth’s atmosphere blocks most infrared light from reaching telescopes on the ground. Webb, with its large mirror, state-of-the-art sensors, and no blurring atmosphere, is expected to see many more exoplanets, including smaller ones down to the size of Saturn or Neptune. Researchers hope Webb’s images will help them understand how and where such planets form.

Next U.K. PM a cipher on science

POLITICS | U.K. scientists expressed unease this week after the country’s Conservative Party chose Liz Truss as the next prime minister. Truss, who served as foreign secretary under outgoing leader Boris Johnson, has said little about research. Her proposed fiscal policy favors innovation, but she has not spoken in favor of Johnson’s goal of doubling government spending on research, and may alter spending plans made by Johnson’s government. Also in question is the United Kingdom’s participation in Horizon Europe, the European Union’s 7-year, €95 billion science funding scheme, given Truss’s hard-line stance on Brexit disagreements with Europe. Unlike some of her fellow Conservatives, she has not denied climate science, but she has not indicated strong support for action to achieve the country’s emissions targets.

Earliest known amputation found

ARCHAEOLOGY | In a cave on the island of Borneo containing some of the earliest known cave art made by our species, scientists have uncovered evidence of a surgical amputation from about 31,000 years ago. The nearly complete skeleton of a *Homo sapiens* individual was found missing the lower half of their left shin and foot. Bony growth fused together the tibia and fibula, indicating the individual, estimated to have been in their early 20s when they died, survived for several years after losing the limb. The clean cut line along the bone, as well as the absence of crush marks that would indicate an animal attack or some other traumatic accident, suggest the lower limb was removed using sharp tools, the researchers report this week in *Nature*.

Chile nixes science-heavy charter

POLITICS | To the dismay of many scientists in Chile, nearly two-thirds of voters on 4 September rejected a draft constitution that would have supported research, environmental policies, and Indigenous rights. The document would have guaranteed scientists freedom of research; it also directed the government to strengthen the development of research and take action on the climate

IMAGE: NASA/ESA/CSA; A. CARTER/UCSC; THE ERS 1386 TEAM; A. PAGAN/STSCI

PHYSIOLOGY

Lung cells give yaks a high-altitude edge

In Tibet's thin, low-oxygen air, 800-kilogram yaks have no trouble sprinting at 40 kilometers per hour. Now, scientists have discovered a special kind of cell in their lungs that may enable these oxlike animals to thrive in the harsh conditions by helping them use oxygen more efficiently. Researchers have long known that certain yak genes and physiological adaptations, such as large hearts and lungs, might increase their fitness at high altitudes. This week in *Nature Communications*, the research team reports finding cells in the lining of yak lung blood vessels with genes—including two linked to high-altitude fitness—that are more active than in similar cells in cattle, which live at lower altitudes.

and biodiversity crises. Last month, more than 1200 scientists endorsed the draft, and both sides of the political spectrum backed many of the science provisions. But the proposal also drew criticism that it would steer Chile too sharply to the left by calling for a more egalitarian society, gender parity, and universal health care.

Gift funds pandemic treatments

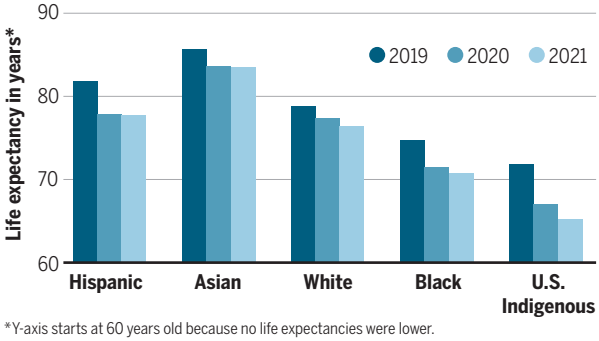
INFECTIOUS DISEASES | Inspired by the record-fast creation of messenger RNA vaccines that blunted COVID-19's impact, a philanthropist is providing AU$250 million ($172 million) over 20 years to bring similar alacrity to the development of therapeutics during future pandemic threats. Geoffrey Cumming, who made a fortune in oil and gas, is putting up what is believed to be the largest gift to medical research in Australia to launch the Cumming Global Center for Pandemic Therapeutics at the University of Melbourne. It aims to create pharmaceutical platforms that could be used to rapidly develop monoclonal antibodies and other antiviral treatments to counter a new pathogen. Some scientists have argued that more lives could have been saved had an effective COVID-19 treatment quickly become available before the vaccines were tested and authorized at the end of that year.

COVID-19 saps U.S. longevity

PUBLIC HEALTH | COVID-19 sent Americans' life expectancy plummeting in 2020 and 2021, in the largest 2-year drop in the past 100 years, the U.S. Centers for Disease Control and Prevention (CDC) said last week. Among racial groups, Native Americans and Alaska Natives saw the biggest decline, from an average of about 72 years in 2019, before the pandemic started, to roughly 65 years in 2021—the same as the overall U.S. life expectancy in 1944. For all Americans, life expectancy in 2021 was just over 76 years, a drop of nearly 3 years from 2019 and the lowest since 1996. Other causes of the decline included accidents, heart disease, and suicide, CDC said.

Indigenous groups see biggest drop

Life expectancy fell sharply for Native Americans and Alaska Natives, despite an active campaign to vaccinate them against COVID-19.

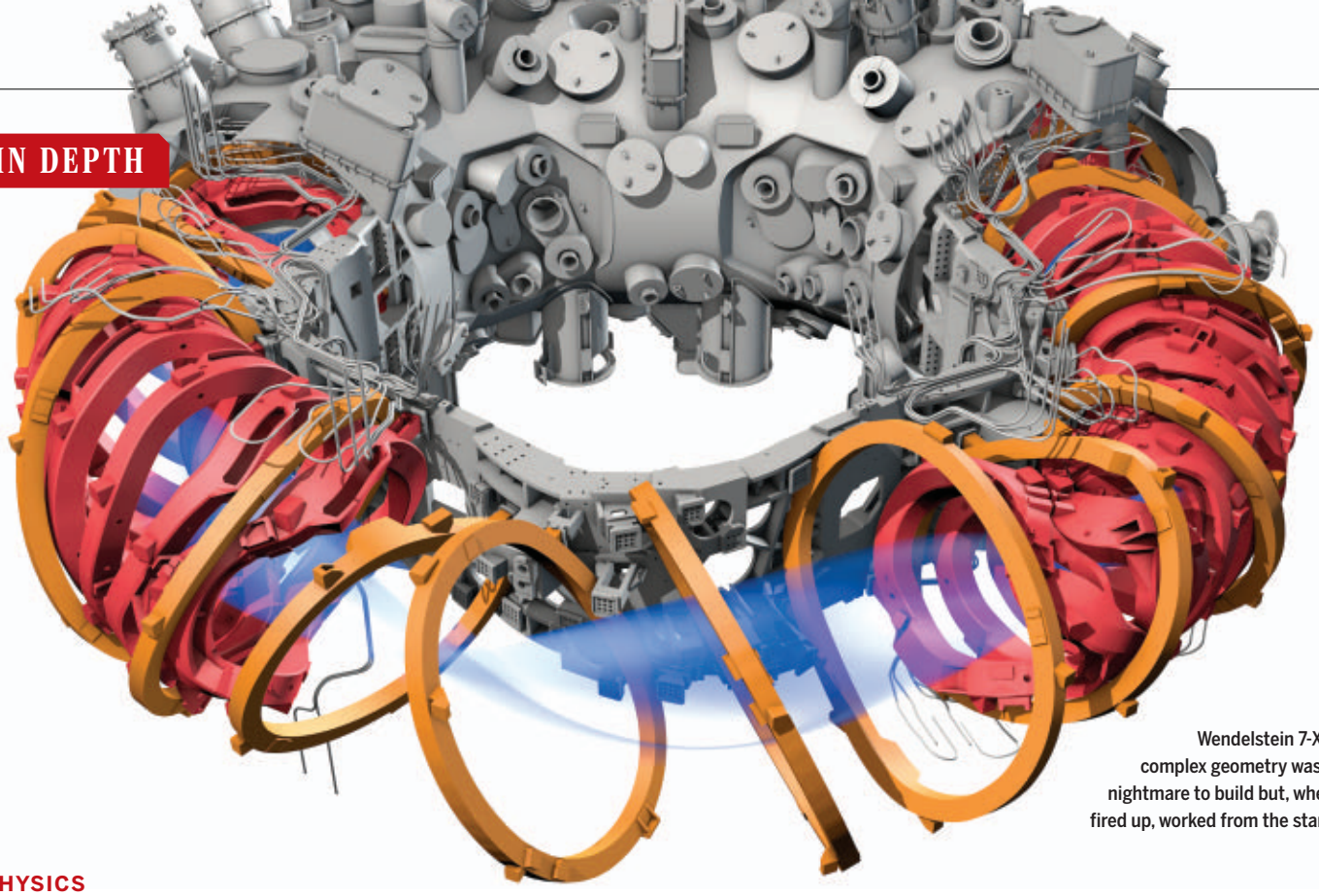


*Y-axis starts at 60 years old because no life expectancies were lower.

La Niña scores a triple

METEOROLOGY | The La Niña weather pattern is likely to last through the end of the year, the first time this century it will have persisted through three Northern Hemisphere winters, the World Meteorological Organization said last week. La Niña is marked by cool surface waters in the eastern tropical Pacific Ocean and is the opposite number to the warming trend known as an El Niño. The effects of La Niña are global: It increases the chances for drought in the U.S. South and the Horn of Africa, which is currently experiencing a severe one. La Niña also tends to deliver more rain to Australia and South Asia—a factor in recent flooding in Pakistan, for example. Climate models suggest La Niña and El Niño events will both become more frequent and intense with global warming.

CREDITS: (PHOTO) WEI LIU; (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) NATIONAL CENTER FOR HEALTH STATISTICS, NATIONAL VITAL STATISTICS SYSTEM, MORTALITY



Wendelstein 7-X's complex geometry was a nightmare to build but, when fired up, worked from the start.

PHYSICS

Twisty device explores alternative path to fusion

Revamped German stellarator should run longer, hotter and compete with tokamaks

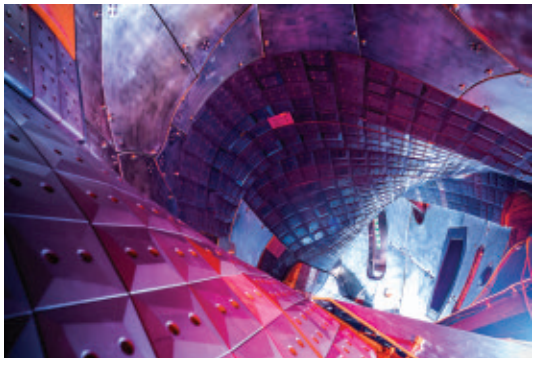
By **Daniel Clery**

Is the search for fusion energy, long dominated by doughnut-shaped devices called tokamaks, about to undergo a shape shift? Just as ITER, the world's largest tokamak—and at tens of billions of dollars the most expensive—nears completion in the hills of southern France, a much smaller testbed with a twistier geometry will start throttling up to full power in Germany.

If the 16-meter-wide device, called a stellarator, can match or outperform similar-size tokamaks, it could cause fusion scientists to rethink the future of their field. Stellarators have several key advantages, including a natural ability to keep the roiling superhot gases they contain stable enough to fuse nuclei and release energy. Even more crucial for a future fusion power plant, they can theoretically just run and run, whereas tokamaks must stop periodically to reset their magnet coils.

In runs of a few seconds, the €1 billion German machine, dubbed Wendelstein 7-X (W7-X), is already getting "tokamak-like performance," says plasma physicist David Gates, proving adept at preventing particles and heat from escaping the superhot gas. If W7-X can achieve long runs, "it will be clearly in the lead," he says. "That's where stellarators shine." Theorist Josefine Proll of the Eindhoven University of Technology is equally enthusiastic: "All of a sudden, stellarators are back in the game." The encouraging prospects are inspiring a clutch of startup companies, including one for which Gates is now leaving Princeton Plasma Physics Laboratory, to develop their own stellarators.

W7-X has been operating since 2015 at the Max Planck Institute for Plasma Physics (IPP) in Greifswald, Germany, but only at relatively low power levels and for short runs. Over the past 3 years, W7-X's creators stripped it down and replaced all the interior walls and fittings with water-cooled versions, opening the way to much longer, hotter runs. At a W7-X board meeting last week, the team reported that the revamped plasma vessel has no leaks and is ready to go. It is expected to restart later this month, on its way to showing whether it can truly get plasma to conditions that, in a future device, would ignite fusion.



Wendelstein 7-X's twisting inner surface is now water cooled, enabling longer runs.

Both stellarators and tokamaks create magnetic cages for gas at more than 100 million degrees Celsius, so hot it would melt any metal container. Heating is provided by microwaves or high energy particle beams. The outlandish temperatures produce a plasma—a roiling mix of separated nuclei and electrons—and cause the nuclei to slam together with such force that they fuse, releasing energy. A fusion power plant would be fueled with a mix of the hydrogen isotopes deuterium and tritium, which react most readily. Research machines like W7-X that aren't trying to generate

CREDITS: (GRAPHIC) IPP, ADAPTED BY C. BICKEL/*SCIENCE*; (PHOTO) IPP/JAN HOSAN

energy avoid radioactive tritium and stick to safer, more plentiful hydrogen or deuterium.

To make their plasma-confining magnetic fields, tokamaks and stellarators employ electromagnetic coils looping around the vessel and through the central hole. But such a field is stronger nearer the hole than the outer edge, causing plasma to drift to the reactor's wall.

Tokamaks tame the drift by making the plasma flow around the ring. That streaming generates another magnetic field, twisting the ionized gas like a candy cane and steadying it. Stellarators use weirdly shaped magnetic coils instead of streaming plasma to produce the twist. The tokamak scheme has long proved the more successful at holding plasma in place, but once plasma physicists had supercomputers powerful enough, they could tweak the complex geometries of stellarator magnets to improve confinement, a process called optimization.

W7-X is the first large, optimized stellarator and contains 50 bizarrely twisted superconducting coils, each weighing 6 tons. Its construction, begun in the mid-1990s, was tortuous, completed 10 years late and costing almost twice the €550 million originally budgeted (*Science*, 23 October 2015, p. 369).

Despite the wait, researchers haven't been disappointed. "The machine worked immediately," says W7-X director Thomas Klinger. "It's a very easy-going machine. [It] just did what we told it to do." This contrasts with tokamaks, which are prone to "instabilities"—the plasma bulging or wobbling in unpredictable ways—or more violent "disruptions," often linked to interrupted plasma flow. Because stellarators don't rely on plasma current, that "removes a whole branch" of instabilities, says IPP theorist Sophia Henneberg.

In early stellarators, the geometry of the magnetic field caused some slower moving particles to follow banana-shaped orbits until they collided with other particles and got knocked out of the plasma, leaching out energy. W7-X's ability to suppress that effect means its "optimization worked as it was supposed to," Gates says.

With this Achilles heel removed, W7-X mostly loses heat through other forms of turbulence—little eddies that push particles toward the wall. Simulating turbulence takes serious computing power, and theorists have only recently got a handle on it. W7-X's upcoming campaign should validate the simulations and test ways to combat turbulence.

The campaign should also showcase a stellarator's ability to run continuously, in contrast to the pulsed operation of a tokamak. W7-X has already operated for runs of 100 seconds—long by tokamak standards—but at relatively low power. Not only were its components uncooled, but the device's microwave and particle heating systems could only deliver 11.5 megawatts of power. The upgrade will boost the heating power by 60%. Running W7-X at high temperature, high plasma density, and for long runs will be the real test of stellarators' potential for producing fusion power. An initial aim, Klinger says, is to get the ion temperature up to 50 million degrees Celsius for 100 seconds. That would put W7-X "among the leading machines in the world," he says. Then, the team will push it for longer, up to 30 minutes. "We'll go step by step, exploring uncharted territory," he says.

W7-X's achievements have prompted venture capitalists to back several startups developing commercial power-producing stellarators. First priority for the startups: Find a simpler way to make the magnets.

Princeton Stellarators, founded this year by Gates and colleagues, has secured $3 million and is aiming to build a demonstration reactor that will forgo the twisted magnet coils of W7-X. Instead, it will rely on a mosaic of about 1000 tiny square coils made of high-temperature superconductor (HTS) on the outside surface of the plasma vessel. By varying the magnetic field produced by each coil, operators will be able to change the shape of the applied field at will. "It takes complexity out of the coils and puts it in the control system," Gates says. The firm hopes to initially develop a reactor that will fuse just cheap, abundant deuterium, to generate not power, but neutrons for manufacturing radioisotopes. If successful, the firm will then aim for a power-producing reactor.

Renaissance Fusion, based in Grenoble, France, has raised €16 million and plans to coat segments of the plasma vessel in a multilayered HTS, forming a uniform coating. Then, using a laser, engineers will burn off tracks within the superconductor to etch a twisting pattern of magnet coils. They aim to make a meter-long test segment over the next 2 years and a full prototype by 2027.

A third firm, Type One Energy in Madison, Wisconsin, received U.S. Department of Energy funding to develop HTS cables with enough bend to be used in stellarator magnets. The company would sculpt pieces of metal with computer-controlled etching machines, carving twisting channels into which the cable is wound to turn it into a coil. "Advanced manufacturing technology opens the door for the stellarator," says cofounder David Anderson of the University of Wisconsin, Madison.

Anderson says the next phase of W7-X's operation will accelerate the boom in stellarator efforts. "With half-hour discharges, you're essentially steady-state," he says. "This is a big deal." ■

FUNDING

New tech law offers billions for research

CHIPS act will fund microelectronics innovation and training through large partnerships

By **Jeffrey Mervis**

Most of the $280 billion in a new law to strengthen the U.S. semiconductor industry and keep the country ahead of China in technology is a 5-year promise, not a reality. But along with the aspirational spending, the recently passed CHIPS and Science Act commits some $13 billion right now for research and training in microelectronics. And U.S. universities are now forming large coalitions with companies and local governments in order to be ready to compete for the money as soon as a trio of federal agencies announces its plans.

"No sane university with a strong interest in microelectronics is sitting this out," says Jesús del Alamo, professor of electrical engineering at the Massachusetts Institute of Technology (MIT), which is involved with several such partnerships.

The act, signed into law on 9 August, funnels $11 billion over 5 years to the U.S. Department of Commerce to create a National Semiconductor Technology Center (NSTC) and a national advanced packaging manufacturing program. Another $2 billion will go to the Department of Defense (DOD) for a microelectronics commons, a national network of university laboratories to develop prototypes for the next generation of semiconductor technologies.

Both initiatives "are aimed at the dearth of lab-to-fab facilities," referring to university laboratories that develop new technologies to be incorporated into semiconductor fabrication plants, or fabs, says Philip Wong, an electrical engineer at Stanford University who heads its nanofabrication center. In addition, the National Science Foundation (NSF) is getting $200 million over 5 years for education and workforce training in microelectronics.

"The [research and training] pipeline is just as important as the fabs, although

it doesn't get nearly as much attention," says Lisa Su, CEO of chipmaker AMD and a member of the President's Council of Advisors on Science and Technology (PCAST). She calls the new law "a once-in-a-generation ... opportunity to fill the pipeline with the next generation of semiconductor technologies."

A PCAST report due out this month will recommend NSTC spend 30% to 50% of its budget on fundamental research across a range of fields in microelectronics, from new materials and energy efficient computing to improved security and health care applications. And "a skilled workforce is a prerequisite for everything," says PCAST member Bill Dally, senior vice president for research at NVIDIA. "We need to do a lot of things to retain our leadership in semiconductors, and all of them require talented people."

New York–based Global Foundries makes high-performance chips.

Two large industry-academic-government partnerships are regarded as the leading contenders for NSTC: the American Semiconductor Innovation Coalition (ASIC), spearheaded by IBM and New York's Albany Nanotech Complex; and the Semiconductor Alliance, which features Intel, Micron, and the MITRE Corporation. Both groups boast a roster of academic heavyweights. MIT, the Georgia Institute of Technology, and Purdue University are members of ASIC, whereas Stanford and the University of California (UC), Berkeley, are closely aligned with the Alliance. Although New York, Virginia, and Texas politicians have proposed their states as hosts, insiders say the center is more likely to be a network of existing facilities spread across the country than a single edifice.

The center's to-do list includes funding multimillion-dollar upgrades to existing fab labs at dozens of universities and providing researchers access to a kind of standard workbench to lower the cost of testing and prototyping new chip technologies. NSTC would also support startup companies that want to commercialize those technologies. In addition, the center would address the need for additional talent at all levels by funding hundreds of new faculty positions, thousands of scholarships, a uniform curriculum in microelectronics with hands-on training, and outreach to middle and high school students.

Only a small number of universities have the capacity to host that training. A white paper issued last year by del Alamo and his colleagues estimates that upgrading a university fab lab to handle 200-millimeter wafers, the size that has become standard in advanced fabs, would cost $80 million. The university would then need $80 million a year in research grants to operate the lab, which could train some 500 graduate students and postdocs. That rules out many institutions, del Alamo says. Even upgrading the select group of universities that remains would still leave the country far short of filling the 42,000 new semiconductor jobs that the CHIPS Act is expected to create, says Tsu-Jae King Liu, UC Berkeley's dean of engineering.

So last year, King Liu spearheaded the formation of the American Semiconductor Academy (ASA), a national network for microelectronics education. She has teamed up with the SEMI Foundation, the nonprofit arm of the industry's trade association, to seek CHIPS funding from DOD and NSF to carry out that vision.

Attracting U.S. students into microelectronics has become a challenge over the past 30 years as semiconductor jobs moved overseas and Google and other U.S.-based companies threw money at recent graduates to write software rather than make devices, says Sanjay Banerjee, a professor of electrical and computer engineering and head of the microelectronic research center at the University of Texas (UT), Austin. UT is part of a consortium that is seeking a slice of the DOD funding to reverse that trend.

Banerjee says the DOD proposal will build on a long-running NSF-funded program, called the National Nanotechnology Coordinated Infrastructure (NNCI), which provides $84 million over 5 years to 16 universities. A DOD grant would be like "putting NNCI on steroids," he says. "It will give us better tools, more staff, and facilities more relevant to industry."

The CHIPS legislation aims to spread research and education opportunities both geographically and by race/ethnicity. That will require participation by more institutions not on the nation's East and West coasts, as well as those serving large numbers of students from groups historically underrepresented in science and engineering. They include the country's historically Black colleges and universities (HBCUs), which as a group rank low on a list of institutions getting federal research dollars.

Consortia vying for the funds must make room for those institutions, says electrical engineer Michel Kornegay of Morgan State University, an HBCU in Baltimore. "Folks talk in general terms about equity and inclusion," says Kornegay, a member of the ASA network. "But they don't spell out how they plan to do that, and whether the people who are going to implement those plans have a track record of success."

The influx of CHIPS money could even widen the gap between the haves and the have-nots if those problems aren't addressed, warns Patricia Mead, dean of engineering at Norfolk State University, an HBCU. "In America, the rich will always get richer," she says. "So you need strong leadership that is going to make a real commitment to broaden the footprint of microelectronics training in this country." ■

PHOTO: ADAM GLANZMAN/BLOOMBERG/GETTY IMAGES

PHOTO: FELIPE DANA/AP IMAGES

Global warming may have already catalyzed irreversible melting of ice sheets in Greenland, a new study finds.

CLIMATE SCIENCE

Warming of 1.5°C carries risk of crossing climate tipping points

Scientists call for concerted effort to forecast points of no return for ice, weather patterns, and ecosystems

By **Cathleen O'Grady**

From melting ice sheets to stressed coral reefs, global warming is changing our world in unmistakable ways. But identifying "tipping points," thresholds past which such transformations become irreversible or self-sustaining, has been more difficult—and controversial. Some researchers argue that emphasizing looming but uncertain points of no return could feed public apathy, rather than energize efforts to curb climate change.

An expansive study of climate tipping points in this issue of *Science* (p. 1171) is likely to fuel that discussion. It synthesizes the most current evidence on how much warming would risk passing 16 tipping points, triggering polar ice collapses, permafrost thawing, monsoon disruptions, and forest and coral reef diebacks. Many of these systems are already stressed by rising temperatures, and the study finds the world might already be within the warming range where the risk is elevated. It also concludes that even under the most ambitious scenario for limiting global warming—to 1.5°C compared with preindustrial levels—the planet could still see dramatic changes.

It's a "timely and thorough piece of work," says Chris Jones, a climate scientist at the Met Office Hadley Centre. The findings are broadly consistent with previous work, he says, but updated and more detailed. He and other climate scientists warn against "cataclysmic" interpretations of the findings. The study also indicates that "a lot of really bad tipping points are still avoidable," says Zeke Hausfather, leader of climate research at the technology firm Stripe.

To estimate tipping points, University of Exeter earth systems scientist David Armstrong McKay and his colleagues gathered evidence from ancient climate records as well as modern observations, model predictions, and current best estimates. They looked at ecological, atmospheric, and other systems to identify those most at risk of abrupt, irreversible, or self-sustaining change as Earth warms. Then, they estimated the minimum amount of warming that might trigger a tipping point in each system, as well as the maximum warming a system might be able to withstand before a catastrophic shift becomes unavoidable. The authors also made a best estimate of where each tipping point sits—somewhere between the extremes—and noted how high their confidence was in each of the 16 forecasts.

Overall, they found that at the current level of global warming—1.1°C since the preindustrial era—Earth has already passed the low-end risk estimate for five tipping points, putting coral reefs, permafrost, and polar ice at risk. Just 0.8°C of warming may have already accelerated the decline of the Greenland Ice Sheet, and as little as 1°C of warming might have put the West Antarctic Ice Sheet on a path toward collapse. "Once [the ice sheet] starts to collapse and loses ice, it actually pushes itself into an even less stable configuration," driving sea level rise around the globe, says climate scientist Nerilie Abram of Australian National University. For both ice sheets, however, the authors estimate that 1.5°C is the more likely tipping threshold, and that the sheets might be able to withstand as much as 3°C of warming without irreversible decline.

Holding global warming to 1.5°C to 2°C—the rough goal of the Paris agreement—could mean exceeding the best estimates for seven tipping points, the study found, causing the loss of mountain glaciers and the disruption of key ocean currents. Although the 1.5°C Paris target was initially a "number of convenience" produced by diplomatic maneuvering, Armstrong McKay says the study reinforces the perils of not meeting that goal.

Other researchers agree. The paper synthesizes a "huge amount of evidence," Abram says, and makes it easier for policymakers and others to see how societal choices could help avoid—or hasten—tipping points.

But too great a focus on specific temperature thresholds could feed arguments that nothing can be done to keep warming to safer levels, warns Bob Kopp, a climate scientist at Rutgers University, Piscataway. "I don't believe in magic numbers in the climate system," he says. "I worry about [tipping points] feeding this idea that there is this threshold, and below it we're OK, above it we're not." Instead, he says, each increment of warming produces additional risk.

The authors hope their work will instigate more concerted research on tipping points. They note, for example, that they have "low confidence" in many of their thresholds.

To develop more certainty, they would like to start comparing tipping point forecasts produced by different climate models. If a range of models produce similar results, "that gives you a lot of confidence," Jones says. If the answers are different, "you know there are still a lot of things we don't understand."

Thomas Stocker, a climate scientist at the University of Bern, thinks the comparison should wait for the next generation of climate models, which are expected to produce even more detailed results. But he and others agree that understanding tipping points is becoming urgent. More scientists, says Tim Lenton, a climate scientist at Exeter and senior author of the study, are "recognizing that some kind of tipping point risk assessment is necessary—and welcome." ■

CHEMISTRY

California EV rules jolt battery science

Move to phase out gas-powered cars will force progress toward faster charging batteries

By **Robert F. Service**

California, known for leading the United States in climate regulations, dropped a bombshell last month: By 2035, the state will ban sales of new gasoline powered cars and light trucks. Most new car sales are expected to shift to battery-powered electric vehicles (EVs). But along with high prices and modest range, current EVs have another big drawback: They are slow to recharge.

Charging stations alone may not be enough to support California's shift to electric vehicles.

Whereas filling a gas tank only takes a few minutes, recharging an EV takes anywhere from the better part of an hour to a day, depending on the charging equipment and the size of the battery.

"There will be a pushback [from car buyers] unless there is a faster charging solution," says Sarah Tolbert, a battery expert at the University of California (UC), Los Angeles. Yi Cui, a materials scientist at Stanford University, agrees. He predicts the broad adoption of EVs will force a revolution in battery design. The need for fast charging, he says, "will definitely provide opportunities for new battery chemistries to emerge." By using new materials for electrodes or charge-carrying ions, he and others have already come up with promising candidates.

Most EVs today use lithium-ion batteries in which one of the two electrodes, the anode, is made of graphite. Graphite has dominated the market because it's cheap, abundant, and able to store enough lithium ions to give cars a range of about 500 kilometers. During charging, the applied voltage pushes electrons into the graphite, attracting lithium ions from the other electrode, the cathode. As the car drives, the lithium lets go of the electrons and travels back to the cathode, while the electrons are routed through the motor, which converts some of their energy into motion, before returning to the cathode.

But graphite anodes are difficult to charge quickly. Most chargers in the United States today use either a standard household voltage of 120 volts (an L1 charger) or 240 volts (L2). Even L2 chargers can require 10 hours or more to fully charge an EV with a typical 500-kilometer range. Still higher voltage L3 chargers, such as Tesla Superchargers, can charge an EV to 80% capacity within 45 minutes. But these nearly 500-volt chargers can cause lithium ions in the graphite to pile up into metal needles called dendrites that can short out the battery and cause it to catch fire. Even if that doesn't happen, high-voltage charging can cause irreversible structural changes in the graphite that shorten the battery's lifetime.

A partial solution may come from simply changing the rates at which graphite-containing batteries are discharged. In a 23 December 2021 *Nature* paper, Cui and his colleagues reported that doubling the discharge rate for the first 2 minutes a battery is in use essentially melts away any built-up lithium dendrites, which can extend a lithium-ion battery's lifetime by 29% and make it stand up better to fast charging.

Another emerging option is to change the anode material altogether. Fifteen years ago, Cui and others showed anodes made from silicon can increase how much charge a battery can store and enable faster charging. Each silicon atom is able to bind four lithium ions, compared with only one for every six carbon atoms in graphite. But pushing so many lithium atoms into a silicon matrix can cause the anode material to swell up to four times in size. And repeatedly charging and discharging the battery typically pulverizes the silicon, killing the battery.

More recently, Cui and others have shown nanoscale modifications to the structure of the silicon, such as forging it into an array of nanowires, can allow the anode to swell and shrink without fracturing, thereby extending the battery life (*Science*, 27 May 2016, p. 1046). Amprius, the company Cui spun out to commercialize the technology, reported in February it has developed a silicon-anode lithium-ion battery with a capacity of 450 watt-hours per kilogram, nearly double that of the 280 Wh/kg cells used in current Tesla EVs. What's more, the new cells can charge to 80% of capacity in just 6 minutes. The company now sells the batteries for drones and other remote aircraft and is working to scale up the technology for EVs.

Other anode materials are also in the works. In 2013, Tolbert, along with UC Los Angeles colleague Bruce Dunn and others, reported that anodes made from the light, gray metal niobium would also enable higher capacity and faster charging than graphite. They processed niobium oxide into a spongelike form, made up of nanoscale tendrils shot through with micron-size pores. This material's very

PHOTO: MARIO TAMA/GETTY IMAGES

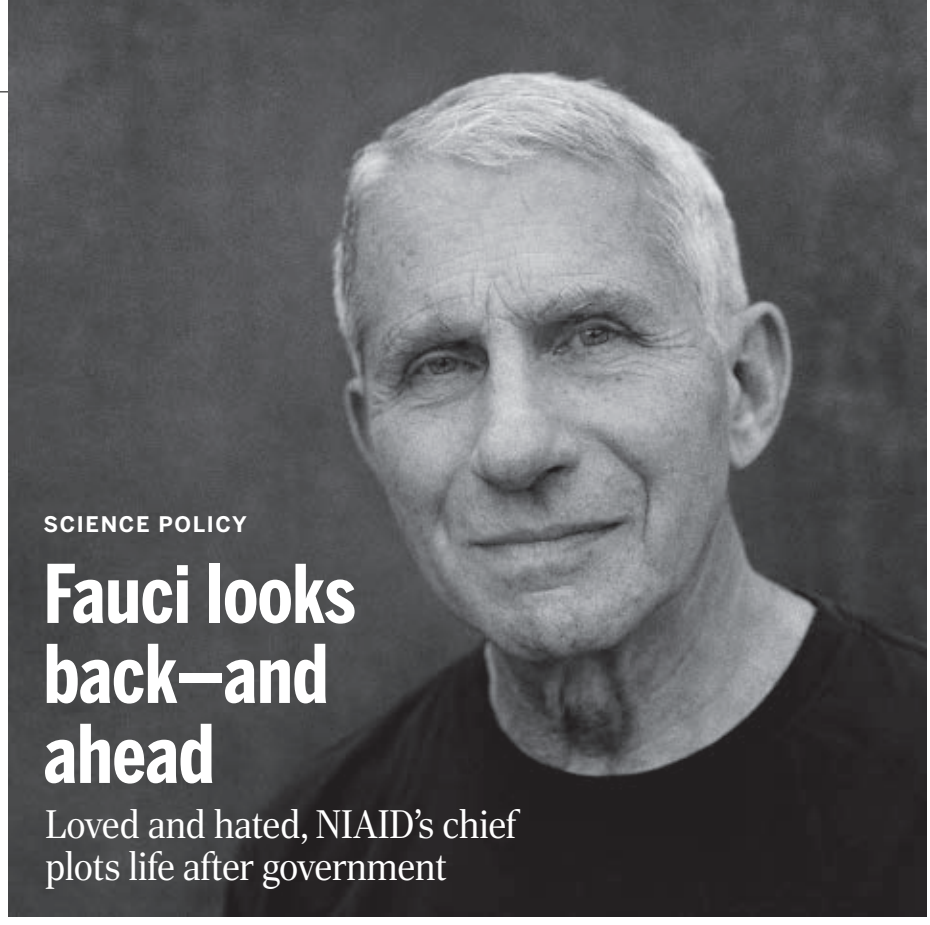
high surface area enables it to hold lots of lithium, and the larger channels enable lithium ions to race through, resulting in faster charging. And unlike silicon, the structure of the niobium-oxide does not change when it grabs and releases lithium ions. Lithium ions nestle close to niobium atoms during charging and simply drift away during discharge, causing less damage to the battery as it goes through repeated charge/discharge cycles.

In 2017, UC Los Angeles licensed its technology to a California startup called Battery Streak. Last month, the company reported it has made palm-size "pouch" cells capable of charging to 80% of capacity in just 10 minutes. (Current EVs use thousands of similar-size cells.) During that fast charging, Battery Streak's cells warm up by just 8°C, compared with graphite-based lithium-ion batteries, which heat by as much as 50°C during high-voltage charging. That should slow battery degradation and extend the life of Battery Streak cells more than 10-fold over current graphite-anode lithium batteries, says Dan Alpern, Battery Streak's vice president of marketing. That increased battery life should offset niobium's price, which is typically more than 30 times that of graphite. Like Amprius, Battery Streak is working to scale up its batteries for EVs.

Replacing the charge-carrying lithium ions with other materials can help as well. In the 24 August issue of *Nature*, for example, Donald Sadoway, a chemist at the Massachusetts Institute of Technology, and his colleagues reported a novel battery design that relies on aluminum ions. Their prototype has a capacity similar to conventional lithium-ion batteries but is capable of recharging in minutes. The battery must operate at near the boiling point of water to allow aluminum ions to move through the device's molten salt electrolyte, which ferries ions between the electrodes. But Sadoway and his team are already working to reduce the operating temperature. If they're successful, the battery could be a blockbuster because aluminum is cheap; compared with lithium batteries, the cost of materials for these batteries would be 85% lower.

Just how all these and other novel battery chemistries may shake out in the marketplace is anyone's guess, says Gil Tal, a transportation technology expert at UC Davis. But he adds it's a safe bet that large-scale EV adoption will cause the battery market to splinter, allowing users to choose their batteries based on whether they prioritize the lowest cost, the fastest charging, the largest capacity, or the longest life. By 2035, Tal says, "The market will be much more diverse." ■

PHOTO: JESSE DITMAR/REDUX



SCIENCE POLICY

Fauci looks back—and ahead

Loved and hated, NIAID's chief plots life after government

By **Jon Cohen**

In 1984, when Anthony Fauci took over as head of the National Institute of Allergy and Infectious Diseases (NIAID), his wife gave him a new plant for his workplace. Both the palm and the 81-year-old physician are still there, the giant plant now crowding the office of one of the most celebrated—and polarizing—scientific figures in U.S. history. But not for much longer. Fauci announced on 22 August that he would step down at the end of the year from both NIAID and his post as the chief medical adviser to President Joe Biden.

"What am I going to do with this plant? It's a monster. I can't fit it in any other place," he joked last week over Zoom. I had seen it before while interviewing the institute chief many times, starting decades ago with the emergence of an earlier pandemic, AIDS.

For many people in the United States, Fauci became the public figure trusted above all others to guide them through COVID-19. The hero worship was evident in Fauci bobbleheads, "In Fauci We Trust" yard signs, and baseball cards that feature him throwing out a first pitch. But many others—including former President Donald Trump and some of his top advisers—turned on Fauci. They saw his advice as inconsistent and misleading, and portrayed him as a threat to the social order, the economy, and the health of the public. In this alternative world, the yard signs say "Fauci for Prison," T-shirts declare "Even My Dog Hates Fauci," and ballcaps call him a fraud or worse. He and his family have faced death threats, and his house has had a Secret Service detail stationed outside for protection.

In a February 2021 webcast interview at the annual meeting of AAAS (publisher of *Science*), Fauci said he wouldn't retire until there were vaccines for HIV, malaria, and tuberculosis (TB). In an interview with *Science* last week, Fauci explained his change of mind, acknowledged some missteps during the pandemic, and discussed what he might do next. The interview has been edited for brevity and clarity. An extended version can be found at https://scim.ag/FauciQA.

Q: What triggered your decision to leave at the end of this year? You had a much longer range view when we spoke at the AAAS meeting, and then you later said you would retire at the end of Biden's first term.

A: No, no, I never said that I would retire at the end of Biden's first term. Please go back and look at my words. Somebody asked me on television, if Donald Trump was the next president, would you stay? And I said, it doesn't matter if he's the next president or if there's another Republican or even if Joe Biden has a second term. I don't plan to be here at the end of the day.

One of the things I had wanted to do deep down was to be here at the end of the

discovery of an HIV vaccine. And this is the truth—and I'm telling you and I haven't told anybody—we're not going to get an HIV vaccine for another decade at least. I was joking when I said malaria and TB.

Q: You really haven't answered the question precisely. Was it the threats to your three daughters and your wife? The attacks in the media and Congress? What finally made you say, I'm done with this job?
A: It was none of the above. As the Trump administration was ending, I was vaguely considering wanting to have at least a few years when I'm still energetic, enthusiastic, and passionate about doing something outside of the realm of the government. Then the president calls me up right after being elected and says, one of the first things I'm going to do, Tony, is ask you to be my chief medical adviser. I said yes, fully thinking that at the end of that first year of the Biden administration, we'd be done with COVID, it really would be in the rearview mirror.

At the end of that year, it became clear that that's not going to happen. And I had long conversations with my best adviser, namely my wife, who said, you know, this is not going to disappear. [Fauci's wife, Christine Grady, heads the bioethics division at the National Institutes of Health's (NIH's) Clinical Center.] So I said to myself, I'm going to be 82 in December, what do I have further to offer?

Despite the knives and the slings and the arrows, I've achieved a degree of being kind of a hero to some people. So let me use that to inspire people. When can I do that? I decided over the spring and summer that the best time to do it is when you're getting somewhat of a steady state with this disease, so I could transition out of here.

Q: Is some of the vitriol toward you about being a flip-flopper with your pandemic advice a result of having to make public health decisions in public in real time?
A: When you're doing an experiment, you collect data, you validate it, you scrub it, you analyze it, and then you write something that's based on science that's not dynamic, but is immutable. When you're dealing with a pandemic response, it's very dynamic, and a somewhat mysterious evolution of a process that has a lot of consequences, because people's lives are involved. The public expects you to analyze the situation and come out with daily proclamations about what should be done. When you're humble and scientific enough to say, you know, we were saying this a week, a month, 2 months ago, but now things have really changed, that's taken as flip-flopping, being wrong, and having made a mistake.

The classic one I know that you're referring to is about masks, right? How many times are we gonna go over that? The surgeon general tweets, please do not buy masks, the Centers for Disease Control and Prevention says, masks are unnecessary. I, as one spokesman, say, you know, we don't know if it works outside of the health care setting. Nobody is fully aware of the aerosol spread. And we had no real idea that 60% of the transmissions were by an asymptomatic person. So, right then, we shouldn't necessarily be wearing a mask. As soon as [the known facts] changed. I said, whoa, wait a minute, we better be wearing a mask in an indoor setting. And that becomes flip-flopping arrogance.

Q: In retrospect, do you think you could have framed it more cautiously in the early days and said, we don't really know enough right now, it doesn't look like we need masks, and there's a shortage of N95s?
A: Yeah, probably so but that's not what people want to hear. They want to hear what should we do? I probably should have couched it in a way that was a little bit more flexible. But I can tell you, Jon, if I did, it's likely I would have been labeled as a wallflower, he didn't know what he was talking about.

Q: So, what are you going to do next?
A: I don't know, Jon, and I'm being very honest with you. It may be hooking up with a university that's willing to make me a senior professor on their faculty. It may be going with a foundation. I can tell you one thing I'm not going to do. I'm not going to sit in my house with a Gmail address.

Q: Looking back, who was the easiest president to work with? I know who the most difficult one was.
A: I wouldn't want to say who was the easiest. They were very different. And the difference often was dictated by the circumstances that defined their presidency. I had a very, very warm, almost grandfather-to-grandson relationship with George H.W. Bush. A fine, fine gentleman. I had probably the most impactful relationship with George W. Bush, because he allowed me to be the chief architect of the President's Emergency Plan for AIDS Relief [PEPFAR]. If you want to look at my career, in lives saved, the most important thing I've done might be PEPFAR. [PEPFAR has invested $100 billion in helping more than 50 countries prevent and treat HIV infections.] But I wouldn't have been in a position to do what I did with PEPFAR unless I had the president trusting me as I did in [Bill] Clinton's administration.

Q: Do you have any advice for NIH about who they should look for to replace you or what type of person?
A: The *type* of person but not a person—that would not be appropriate. You need somebody who clearly has scientific credibility. It also has to be somebody who is articulate enough to navigate the degree of public exposure you will have, because you will be involved in the next public health crisis, which very often is an infectious disease.

Q: What do you want to accomplish between now and December?
A: I'd like to guide the good research being done in the arena of both mucosal, nasal vaccines for respiratory diseases, as well as more durable vaccines that protect against entire families of viruses. I'm a pusher. We meet as a group at least three, four times a week and I go OK, what are we doing? I try to be not a pain in the ass, but I'd like to keep my foot on the pedal between now and then to keep that going.

Q: I've seen loads of people confront you over more than 30 years. No one seems to get under your skin as much as Senator Rand Paul (R–KY). Why does that guy piss you off so much? [Paul has accused Fauci of helping trigger the pandemic by funding scientists in Wuhan, China, who, Paul and others allege, either created SARS-CoV-2 or had collected a bat virus that accidentally leaked from their lab.]
A: I came into what I thought was a good faith, oversight hearing where politicians ask questions for the purpose of improving the situation, for the purpose of protecting and preserving the health of the country. When you start off by saying, how do you explain the fact that you are responsible for the death of 5 million people? No way am I going to stand for that on public television, on C-SPAN. Sorry, Jon. No way.

Q: If Republicans gain the majority in the House of Representatives and they hold hearings, will you testify?
A: If I'm asked to testify, I'll testify. I have nothing to hide. I can explain everything I've done and I could defend everything I've done. But if it becomes clear that it's not an in-good-faith oversight, but a character assassination, I might not play ball.

Q: A last question for you. Do you have a motto that you told your kids, like, this is what I learned in life?
A: Yeah, it's what I go by. It's called precision of thought and economy of expression. Know what your message is, know your audience, and say it in as few words as possible. ■

In Magude district in southern Mozambique, Zaidina Antonio Zwane gives her 2-year-old son Ernesto a glass of water to wash down his malaria treatment.

PHOTO: TOMMY TRENCHARD/PANOS PICTURES/REDUX

LINGERING FEVER

The battle against malaria in Africa has stalled. Can research in Mozambique explain why—and how to get it back on track?

By **Leslie Roberts**

Moisés Mapanga, a burly man of 49, is the bait. At 6 p.m. on a mid-April evening, he climbs into an orange tent outside his one-room house in Matutuíne, a hot, swampy district near Maputo, the capital of Mozambique. He will remain there for the next 12 hours, except for a 30-minute dinner and bathroom break.

Mosquitoes hunting for their next blood meal fly into the tent, where they are attracted not only to Mapanga, who sleeps under a second tent to protect him from their bites, but to a red light atop a lantern-shaped mosquito trap. Instead of dining, the insects are sucked into a collection cup by a battery-powered fan. The next night Mapanga will repeat the drill with the tent inside his house.

Junior entomologist Mara Máquina of the Manhiça Health Research Centre (CISM) and her colleagues are on site to inspect the catch. They are testing house after house in Matutuíne to figure out which species of malaria-carrying *Anopheles* mosquitoes occur there and whether they are biting indoors or outside. The answers might help reduce the disease's horrendous toll in Mozambique.

But the project also aims to help answer a bigger question: Why, after Mozambique and its international partners have spent years pouring money and energy into fighting malaria, have cases plateaued, and in some places increased? "We do almost everything we can to decrease the burden, but we are not seeing the impact we would expect," says Pedro Aide, malaria area coordinator at CISM.

The pattern is similar in many countries in southern Africa. After more than a decade of striking successes against the disease, cases are now bumping along a generally flat line, slightly up one year, down the next, with no real progress. Researchers hope Mozambique, which has the fourth-largest malaria burden in the world, can shed light on what is happening across the region.

The question bothers donors as well. Last year, the Global Fund to Fight AIDS, Tuber-

culosis and Malaria, which in late 2020 gave Mozambique $200 million for 2.5 years to control malaria, requested an analysis to figure out why it is not seeing a greater payoff. Unfortunately, "There's not one obvious answer," says James Colborn, a senior malaria adviser at the Clinton Health Access Initiative who led the analysis.

Part of the problem is that "the mosquitoes are fighting back," Aide says. Tried-and-true interventions against malaria, such as bed nets treated with long-lasting insecticide and spraying the indoor walls of houses with insecticides, are losing power because mosquitoes are changing their behavior and resistance to the chemicals has become rampant. The vagaries of human behavior factor in as well, as does poverty, with which malaria is intertwined, and a lack of health services. And a warmer, wetter climate, on top of the cyclones and flooding that pound the country almost every year, is creating even better conditions for mosquitoes to flourish.

Researchers at the Manhiça Health Research Centre collect mosquitoes by placing a trap inside a tent.

Whatever the causes, it's clear that fixing the problem will require new tools—especially in central and northern Mozambique, where malaria still kills tens of thousands of children each year.

GLOBAL SPENDING on malaria control and elimination started to go up sharply after the turn of the century, thanks mostly to the launch of the Global Fund in 2002 and the U.S. President's Malaria Initiative (PMI) in 2005. Malaria spending surpassed $3 billion annually by 2010 and has plateaued since then. As spending elsewhere began to drop, however, support for Africa, which has by far the world's worst malaria problem, continued to rise. It passed $2.5 billion in 2020.

The results were stunning—at least initially. Between 2005 and 2015, the number of estimated yearly cases globally fell by 30% and the mortality rate dropped by 47%. But then progress ground to a halt at what the World Health Organization (WHO) calls "unacceptably high" levels. There were an estimated 241 million cases and 627,000 deaths in 2020, mostly in children in Africa.

The paradox is acute in Mozambique, whose entire population of 31 million is at risk. PMI has so far spent more than $415 million to fight malaria in the country and the Global Fund $530 million. Yet, countrywide incidence jumped from 226 cases per 1000 residents per year in 2015 to 368 per 1000 in 2020, an increase of more than 60%, according to PMI. More than 11 million cases were diagnosed in 2020, about 90% of them caused by *Plasmodium falciparum*, the most dangerous of the four malaria parasite species. Another measure of malaria intensity is prevalence, the percentage of children under age 5 who have malaria parasites in their blood, regardless of whether they are sick. In Mozambique, that number hasn't budged from about 40% nationwide since 2011.

One encouraging sign is that the number of deaths has been declining gradually in Mozambique, thanks to better access to rapid diagnostic tests and the best drugs, artemisinin-based combination therapies (ACTs). But the country still saw an estimated 23,700 deaths in 2020, according to PMI. Malaria is a leading cause of death among young children.

The toll is not evenly distributed. Mozambique stretches 2500 kilometers along the Indian Ocean, and malaria transmission varies dramatically with climate and topography. It's highest—and has increased sharply—in the country's center and the north, with their humid tropical climates. Transmission is much lower, and has decreased, in the drier lowlands of the south. The stark difference is reflected in the prevalence rates in children under age 5, which range from 57% in Cabo Delgado, a province in the far north, to 1.3% in Maputo, the country's southernmost province (see map, p. 1141).

Poverty exacerbates malaria—and Mozambique is one of the world's poorest countries. The underprivileged live in flimsy housing that offers little protection from mosquitoes, and they are the least likely to receive care. Even when malaria does not kill, the disease perpetuates poverty because the sick cannot work or go to school.

MAPUTO PROVINCE is Mozambique's malaria success story—but even there, efforts have fallen short of their goal. In the 1990s the government began to fight malaria in earnest, in large part to assuage its neighbors, South Africa and Eswatini, formerly Swaziland. Those countries wanted to wipe out malaria altogether and were frustrated that migrant workers and others from Mozambique kept bringing it back in. "They look to us and say, 'Hey, you guys need to fix your problem there,'" says Abuchahama Saifodine, the U.S. Agency for International Development's Maputo-based resident adviser. The Mozambican government also hoped getting rid of malaria would foster development and tourism.

The south has been a testing ground for new techniques ever since. In 2015, with funding from the Bill & Melinda Gates Foundation and the La Caixa Foundation, the government and its partners, CISM and the Barcelona Institute for Global Health, embarked on a bold, 2-year pilot to test the feasibility of eliminating malaria in Magude, one of Maputo's seven districts.

They threw everything they had at the problem. All residents received insecticide-treated bed nets. Pregnant women, who are at risk of severe disease and anemia, were offered prophylactic antimalarials when they sought antenatal care. In addition, every house received indoor spraying once a year to drive down the mosquito population. The project also tried something new for Mozambique: providing the district's entire popu-

PHOTO: CYNTHIA R. MATONHODZE

lation with an antimalarial twice a year, regardless of whether they were sick. The hope was that this mass drug administration would not only cure the ill, but also clear the parasite from those who were asymptomatically infected.

During the study, prevalence in Magude fell from 9.1% to 1.4%—but not to zero. Malariologists attribute this to residual malaria transmission—the transmission that continues even when all available mosquito control tools are deployed effectively. All over the country, "residual malaria transmission is a big issue and a really big headache," says Nelson Cuamba, an entomologist at the Ministry of Health's National Malaria Control Programme (NMCP).

The human-baited tent traps point to one cause. At the first house the CISM team visited in mid-April, the trap came up empty when the volunteer slept in a tent inside the house, except for a few *Culex* mosquitoes, which don't transmit malaria.

But when Mapanga spent the night outside in his tent, the traps snagged a few *Anopheles* mosquitoes, which Máquina can readily identify by the black-and-white spots on their wings. That means the mosquitoes were biting outside, where bed nets and indoor spraying offer no protection.

The findings reflect the intense, constant pressure on mosquitoes to adapt to control measures, which has resulted not only in insecticide resistance, but also in behavioral changes. Mosquitoes are biting more outdoors—sometimes right outside the door—or earlier in the evening or later in the morning, when people are not protected by nets. Or they may feed indoors but rest outdoors, safe from the insecticides.

In a 2019 paper, researchers from Imperial College London reported that outdoor biting among three vector species across sub-Saharan Africa may have increased 10% since 2000. That's a huge problem because there are no good tools to thwart outside biters, says Baltazar Candrinho, who heads NMCP. "It is happening everywhere," he says.

Another experimental setup is yielding more clues about the behavior shift. Entomologists lay a white sheet on the floor in the morning, spray the house with a commercial insecticide, "and see what falls down dead," says Krijn Paaijmans of Arizona State University, Tempe, who coordinates the entomology group at CISM. Window exit traps, which allow mosquitoes to fly in but not out, snag the ones attempting to leave the house.

Such studies have shown the balance within the *Anopheles* genus is changing. *An. funestus*, which bites and rests indoors, used to be the most abundant vector in Maputo. Now, *An. arabiensis*, which can bite indoors or out, is the most plentiful. In Gaza and Inhambane provinces, where indoor spraying has not been used, the team still finds lots of *An. funestus*.

There are other reasons why mosquito control alone may not be enough to end malaria. Bed nets work very well—when people are sleeping under them. But in many rural communities, people stay outside late, cooking, socializing, or doing school work and avoiding hot houses made even steamier by metal roofs, Paaijmans and his colleagues have found. After people go indoors, they don't necessarily jump in bed, leaving them vulnerable to mosquitoes.

People don't always welcome indoor spraying, either. It's inconvenient—families have to move their belongings outside—and smells bad, and some people don't trust the army of sprayers or the government that sends them, Saifodine says. After they leave, some residents wash or replaster their walls or build new rooms, Paaijmans's team has found.

IN RECENT YEARS, Mozambique has shifted its antimalaria focus to the center and the north of the country. The move was part of an international push to concentrate on the high-intensity regions, where the battle is much harder but more lives can be saved. "We want to put our resources in the highest burden places where we can get the most impact in a short time," Candrinho says. The Global Fund has stepped up its funding to pay for the intensified attack, and PMI spends its money in this part of the country. But progress has not only stalled; it has reversed.

There's too much malaria in the north and the center to try to chase it from the population with mass drug administration or by using the finely honed tools deployed in Magude, such as following every person diagnosed with malaria home and giving family members antimalarials to prevent further spread. "If you are in Zambezia [province] and there are 2 million cases, there is no way you can follow patients home," Saifodine says.

Rather, NMCP focuses on the basics. In 2016, Mozambique began to strive for universal bed net coverage, which means one net for every two people. It also tried to ensure health facilities are stocked with rapid diagnostic tests and ACTs and that every-

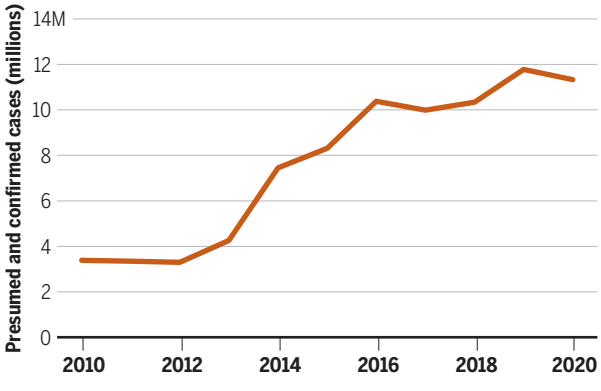
A stubborn enemy

In Mozambique's center and north, where malaria prevalence is highest, cases have gone up in recent years. In Magude district, in the south, an all-out attempt to eliminate the disease fell short.

Prevalence in children aged 6–59 months (% in 2018)

1–15 16–29 30–43 44–57

Mozambique's rising malaria toll



CREDITS: (MAP) K. FRANKLIN/*SCIENCE*; (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) U.S. PRESIDENT'S MALARIA INITIATIVE *MOZAMBIQUE MALARIA OPERATIONAL PLAN FY 2022*; WORLD HEALTH ORGANIZATION *WORLD MALARIA REPORT 2021*

Celso Melembe, a technician at the Manhiça Health Research Centre, uses a filtered tube to transfer mosquitoes in a study into the efficacy of insecticide-treated bed nets.

one who shows up with a fever is tested for malaria and treated. Also, all pregnant women who seek antenatal care are offered chemoprevention.

So why did cases continue to surge? The study Colborn and colleagues just finished for the Global Fund mentioned several possible reasons. One somewhat hopeful explanation is that Mozambique has gotten better at detecting and testing cases, Colborn says, which could explain part of the frustrating statistics. But Colborn's analysis listed many other potential factors. Insecticide resistance in mosquitoes may be worse than researchers realize. The shifts in mosquito behavior that Paaijmans's team is documenting may be having a profound impact across the country. "If we were able to measure all potential factors affecting transmission, we would probably have a good idea of why we see stagnation or an increase in malaria," Colborn says. At this stage, "we can say: This is what we don't know and what we need to know."

There are also stubborn realities in this vast, very poor country. "Universal coverage is not universal," Candrinho concedes. The most recent national data, from 2018, showed just 69% of the population had access to a bed net, despite campaigns to distribute 16 million to 17 million nets across the country every 2.5 years, about the life span of a net. Across central and northern Mozambique, there are often no community health workers. People have to walk 2 to 3 hours to get to a health facility, and many don't. Many women don't seek antenatal care.

Once they do get to a clinic, a number of things can go wrong. Health workers don't always ask whether the patient has a fever; of those with fever, only some are tested for malaria, and of those, only some receive ACTs, a recent study by Candrinho, Colborn, and colleagues found.

A big part of the problem in central and northern Mozambique is that malaria is so common people tend to shrug it off. Some people don't know malaria is transmitted by mosquitoes or how to protect themselves.

That's why on a hot spring day a religious leader in Nampula province, Juma Alberto Munlela, is talking about malaria with a group of about 15 men in a straight row of plastic chairs under the span of a huge tree. They live in Namaita, a devastatingly poor community far off a paved road in Rapale district.

Munlela is with the Inter-Religious Program against Malaria (PIRCOM). Funded by PMI, it employs respected religious leaders of all faiths—Muslims, Christians, Hindus, and Baha'i—to educate people about malaria, and is spreading its message in five of Mozambique's 11 provinces.

Munlela begins with a prayer before asking whether anyone knows where malaria comes from. Mosquitoes, a man volunteers, and they like standing water, so it is important to clean your house. Malaria gives you fever, headaches, and convulsions, another adds. Malaria is deadly, Munlela explains, and children and women are especially vulnerable. If anyone has a fever, go to a health facility right away and get tested for malaria. Take the antimalarial for the full 3 days even if you feel better—if you take it just 2 days malaria is not over. "God wants you to go to the hospital," he intones.

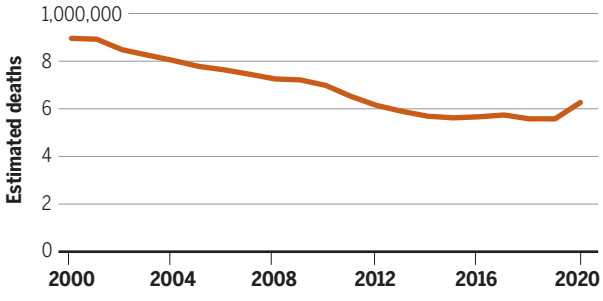
Next, Munlela stresses the importance of sleeping under a bed net, and his PIRCOM colleagues demonstrate how to assemble a bright blue one. A young man follows their example, climbs in, and laughingly mimes sleep. But there's a problem, another man says: The community has no bed nets.

PIRCOM staff will return to Namaita in a few weeks to see whether the community is following their leader's advice, for instance by seeking care when they feel ill. "Changing behavior is not something you say today and it happens tomorrow," says Herminio Rafael Guitutela, PIRCOM's provincial coordinator in Nampula. "It's a process."

AS RESEARCHERS SEARCH for answers to Mozambique's malaria puzzle, one

Success, followed by stagnation

An intensified battle against malaria that started around 2000 caused worldwide deaths to drop—until progress stagnated about a decade ago.



CREDITS: (PHOTO) CYNTHIA R. MATONHODZE; (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) WORLD HEALTH ORGANIZATION *WORLD MALARIA REPORT 2021*

thing is abundantly clear: New strategies and tools are needed if the country is to bring down its sky-high burden.

To thwart insecticide resistance, NMCP is rolling out next-generation bed nets, which use a combination of two insecticides or another formulation, in four provinces. Colborn thinks that is one reason cases dipped in those provinces last year. "It is the first obvious success I have seen from one intervention." Next-generation nets will be distributed to the rest of the country over about the next 6 months, he says.

Last year NMCP, supported by the Global Fund, tried mass drug administration in an entirely different context—not for malaria elimination in the south, but to reduce disease and death in Cabo Delgado, the northernmost province, where conflict has displaced hundreds of thousands of people and caused malaria to spike. The drugs were given in two areas where most of the displaced are concentrated. Data are still being analyzed, but the strategy helped, Aide says: "Malaria transmission continued, but we didn't see the expected increase in cases after the peak malaria season." NMCP may try the approach in other high-burden provinces as well.

The nonprofit Malaria Consortium, in collaboration with NMCP and CISM, is conducting a 2-year pilot study of a brand-new strategy for Mozambique: seasonal malaria chemoprevention, which involves giving all eligible children ages 3 months to 59 months intermittent doses of a powerful antimalarial drug combination—|sulfadoxine-pyrimethamine and amodiaquine (SPAQ)—during the high-transmission season.

Seasonal chemoprevention is widely used in the Sahel, where it's been shown to reduce malaria incidence by 75%. But it has never been tried in sub-Saharan Africa, for a couple of reasons. Malaria is highly seasonal in the Sahel, Aide says—it is either on or off—so it is very clear when children are most vulnerable. In Mozambique, malaria is seasonal as well, but the incidence never drops to zero. Also, resistance to sulfadoxine-pyrimethamine among malaria parasites is high in Mozambique, which could blunt SPAQ's effectiveness.

But the malaria burden remains so stubbornly high that in 2020 NMCP decided to give the strategy a try in two districts in Nampula province. The first phase, last year, showed seasonal chemoprevention is safe, feasible, and widely accepted, says Maria Rodrigues, Mozambique country director of Malaria Consortium. Now, the collaborators are testing its effectiveness.

In one of the districts, remote and dusty Malema, local women recruited for the study were going door to door in April, talking to mothers and distributing SPAQ. Community leaders had already paved the way, explaining that the strategy is prevention, not treatment, and helping dispel any rumors.

At the first house, a 2-year-old girl with huge brown eyes, perched on her father's hip, is feverish and subdued. Because she is sick, she can't take SPAQ, Rodrigues explains. Instead, the team urges the father to take her to the health center immediately to be tested for malaria.

At other houses, the community distributors show mothers how to dissolve the tablet, making an orange-flavored liquid that most children like. For those too young to drink from a cup, mothers make a paste and spoon it into the child's mouth. When the distributors are done, they mark the door with chalk, recording the date, how many children received the drug, and whether they need to come back to catch any they missed. Several days later they will return and collect a card that the mother has filled out showing how many doses she delivered.

So far, the Malaria Consortium has treated about 114,000 children. Preliminary data are very encouraging, Rodrigues says: A comparison with two control districts suggests seasonal chemoprevention reduced malaria incidence about 85%. This year, NMCP will expand the strategy across all of Nampula province, targeting 1.3 million children.

Mozambique is also one of two countries, with Tanzania, testing a novel use of ivermectin, a drug used to fight parasites in people and animals that also kills mosquitoes feeding on people who have taken it. In the study, led by the Barcelona Institute for Global Health, researchers are giving ivermectin to both humans and their livestock, which mosquitoes also feed on, in mass campaigns for two consecutive years. Used in combination with other tools, "it could be a great innovation," Saifodine says.

Yet another tool may be coming. Mozambique was one of the early trial sites for RTS,S, the first-ever malaria vaccine, which WHO approved last year for widespread use in areas of moderate to high malaria transmission. The vaccine is in short supply and far from perfect, with an efficacy of just 30%. But if you consider the number of people at risk in Mozambique, it could make a big difference, Candrinho says. NMCP and its partners are considering how it might be used in combination with other tools and whether to apply for funding.

PHOTO: NUNO MARIO/MALARIA CONSORTIUM

Health workers in Nampula province mark a house where they have given children prophylactic drugs during peak malaria season. Early results suggest seasonal chemoprevention sharply reduces incidence.

Like other innovations, the vaccine costs money, which, despite the influx of external funding, is scarce in Mozambique. None is a panacea. But Saifodine remains optimistic that, working with the tools they have and trying to figure out where each makes sense, Mozambique can finally start to make progress. "If we can find a solution for funding, a lot of great things can happen," he says. ■

Leslie Roberts is a science journalist in Washington, D.C. This story was supported by the Pulitzer Center.

INSIGHTS

Although children generally develop mild COVID-19, some can develop severe disease, and so it is important they are protected.

PERSPECTIVES

VIEWPOINTS: COVID-19

COVID-19 and children

There has been substantial research on adult COVID-19 and how to treat it. But how do severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections afflict children? The COVID-19 pandemic has yielded many surprises, not least that children generally develop less severe disease than older adults, which is unusual for a respiratory disease. However, some children can develop serious complications from COVID-19, such as multisystem inflammatory syndrome in children (MIS-C) and Long Covid, even after mild or asymptomatic COVID-19. Why this occurs in some and not others is an important question.

Moreover, when children do contract COVID-19, understanding their role in transmission, especially in schools and at home, is crucial to ensuring effective mitigation measures. Therefore, in addition to nonpharmaceutical interventions, such as improved ventilation, there is a strong case to vaccinate children so as to reduce possible long-term effects from infection and to decrease transmission. But questions remain about whether vaccination might skew immune responses to variants in the long term. As the experts discuss below, more is being learned about these important issues, but much more research is needed to understand the long-term effects of COVID-19 in children. **–Gemma K. Alderton**

PHOTO: TYRONE SIU/REUTERS

Why is COVID-19 generally milder in children?

By **Carl A. Pierce[1], Kevan C. Herold[2], Betsy C. Herold[1,3]**

Why infants and young children have milder clinical courses with COVID-19, accounting for ~0.1% of deaths, but are more vulnerable to established pathogens has been a key question since the onset of the pandemic. This may be attributed to differences in susceptibility. All segments of the population were naïve to SARS-CoV-2, whereas older children and adults have protective immunity to established viruses because of prior exposures and vaccines. Although the underrepresentation of children in severe COVID-19 cases was initially hypothesized to reflect decreased susceptibility, no age-associated difference in household transmission, expression of angiotensin-converting enzyme 2 (ACE2, the receptor for viral cell entry), antibody titers against other coronaviruses, or SARS-CoV-2 viral loads in nasopharyngeal swabs have been detected (*1*–*4*). These findings, coupled with immunological studies, suggest that the more favorable outcomes likely reflect rapid engagement of the pediatric mucosal immune system.

Studies of immune responses in hospitalized children and adults during the first wave of the COVID-19 pandemic found that pediatric patients had less robust memory T cell responses and lower neutralizing and Fcγ-receptor activating antibody responses than adults (*2, 5*). Coupled with data showing an age-dependent decrease in serum levels of the cytokines interferon-γ and interleukin-17, these findings suggested that children may mount a more vigorous innate (pathogen nonspecific) response that promotes viral clearance and precludes the robust adaptive immune response to SARS-CoV-2 that contributes to severe disease in adults (*2*). This hypothesis was strengthened by direct study of mucosal responses at the time of COVID-19 diagnosis: RNA sequencing data and measurement of cytokines demonstrated a brisk innate response in the nasal mucosa of children versus adults (*3*). These findings may be a consequence of trained immunity resulting from more frequent respiratory infections in children, leading to higher baseline innate activity in pediatric relative to adult nasal mucosae (*6, 7*).

Another hypothesis is that the naïve character of the pediatric adaptive immune system contributes to milder outcomes. Analysis of peripheral blood immune cells in pediatric and adult COVID-19 patients showed an increased frequency of naïve T cells, depletion of natural killer (NK) cells, and a lower frequency of cytotoxic T cells in children compared with adults (*6*). Likewise, the T cell receptor repertoire from young children who seroconverted after COVID-19 showed reduced clonal expansion compared with adults (*8*). The skewing away from cytotoxic activity may avert an overexuberant adaptive immune response that contributes to hyperinflammation, which is characteristic of severe disease.

Clinical experiences with newer SARS-CoV-2 variants, which are potentially more transmissible, suggest that the innate response continues to play an important role, although there is concern that future variants with greater capacity to evade innate immunity may emerge. Thus, vaccinations, which are also approved for young children, are critical as SARS-CoV-2 transitions from a pandemic to an endemic virus. Notably, vaccines also protect against MIS-C, a rare but potentially severe post–COVID-19 pediatric sequela. The findings underscore that differences in outcomes with new and old infections depend on the host, their prior infectious history and immune status, and the pathogen(s) involved. Studies across the age spectrum are needed to identify these contributing factors and guide public health policies.

Why do some children develop MIS-C?

By **Janet Chou[4,5]** *and* **Adrienne Randolph[4,6,7]**

An obese but otherwise healthy 9-year-old boy develops fevers, rashes, vomiting, neutrophilia, lymphopenia, elevated inflammatory markers, and coagulopathy. He had asymptomatic COVID-19 4 weeks before displaying these symptoms. Owing to myocarditis and hypotension, he receives immunosuppressive dosing of intravenous immunoglobulin (IVIG) and glucocorticoids, enabling recovery in 1 week. This case is representative of MIS-C (*9, 10*), defined by fever, inflammation, and involvement of at least two organ systems and requiring hospitalization. Coronary artery aneurysms occur in over 8% of patients (*11*). Unlike acute COVID-19, MIS-C is a postinfectious syndrome occurring within 3 to 6 weeks of mild or asymptomatic SARS-CoV-2 infection (*9*). Although severe COVID-19 occurs most frequently in the elderly, healthy children between 6 and 12 years of age most commonly develop MIS-C. In 2020, MIS-C occurred in ~3 in 10,000 US persons under 21 years, with a mortality rate of ~0.8%. Obesity is present in ~30% of children with MIS-C (*12*), but its role and what triggers MIS-C remain unknown. The incidence and severity of MIS-C are lower after infections with the Omicron variant, compared to the Alpha or Delta variants. This has been attributed to differences in inflammatory responses associated with each variant and enhanced host immunity after COVID-19 vaccinations or SARS-CoV-2 reinfections (*13, 14*).

Patients with MIS-C have activated neutrophils, monocytes, T cells, B cells, NK cells, and dendritic cells (*15, 16*). Increased circulating chemokines and cytokines facilitate homing of these activated immune cells to inflamed organs, most commonly the gastrointestinal and cardiovascular systems. Autoantibodies occur in some patients with MIS-C, but it is unknown if they drive autoimmunity or are byproducts of a hyperinflamed state. The SARS-CoV-2 spike protein has structural similarities to superantigens that can nonspecifically activate T cells (*12*). The expansion of T cells with specific T cell receptor Vβ repertoires, a characteristic of superantigen activation, has been observed in some patients with MIS-C (*17*).

Although effective treatments for MIS-C are available, the causes of this postinfectious syndrome remain unclear. It is challenging to obtain biospecimens from critically ill children before treatment,

[1]Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA. [2]Departments of Immunobiology and of Medicine, Yale School of Medicine, New Haven, CT, USA. [3]Department of Pediatrics, Albert Einstein College of Medicine, Children's Hospital at Montefiore, Bronx, NY, USA. betsy.herold@einsteinmed.edu; kevan.herold@yale.edu [4]Division of Immunology, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA. [5]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [6]Department of Anesthesiology, Critical Care, and Pain Medicine, Boston Children's Hospital, Boston, MA, USA. [7]Department of Anaesthesia, Harvard Medical School, Boston, MA, USA. janet.chou@childrens.harvard.edu [8]Manchester University Foundation Trust and School of Biological Sciences, University of Manchester, Manchester, UK. [9]Long Covid Kids & Friends Charity, Crowhurst, UK. binita.kane@mft.nhs.uk [10]William Harvey Research Institute, Queen Mary University of London, London, UK. [11]Clinical Operational Research Unit, University College London, London, UK. d.gurdasani@qmul.ac.uk; c.pagel@ucl.ac.uk [12]Texas Children's Hospital Center for Vaccine Development, Departments of Pediatrics and Molecular Virology and Microbiology, National School of Tropical Medicine, Baylor College of Medicine, Houston, TX, USA. [13]Department of Biology, Baylor University, Waco, TX, USA. [14]Hagler Institute for Advanced Study, Texas A&M University, College Station, TX, USA. [15]Scowcroft Institute of International Affairs, Texas A&M University, College Station, TX, USA. [16]James A. Baker III Institute for Public Policy, Rice University, Houston, TX, USA. [17]School of Public Health, University of Texas, Houston, TX, USA. hotez@bcm.edu [18]Department of Ecology and Evolution, University of Chicago, Illinois, USA. [19]Department of Microbiology, University of Pennsylvania, Philadelphia, PA, USA. cobey@uchicago.edu; hensley@pennmedicine.upenn.edu

and there are no animal models of this disease. The hypothesis of persistent, low-level SARS-CoV-2 replication driving MIS-C is based on one study finding circulating SARS-CoV-2 spike protein in a subset of patients with MIS-C (*12*), although this was not replicated (*18*). Similarly, extremely rare cases of MIS-C occurring after COVID-19 vaccination prompted concerns that vaccine-derived spike protein could be a trigger (*19*). However, studies have shown that COVID-19 vaccination reduces the risk of MIS-C (*20*), possibly by reducing the risk of SARS-CoV-2 infection, and the overall incidence of MIS-C is decreasing. Drivers of hyperinflammation, including genetic variants impairing restraint of immune activation, have been identified as additional risk factors (*12*). Future studies need to resolve this paradox of a hyperinflammatory disease affecting previously healthy children with antecedent asymptomatic or mild SARS-CoV-2 infections. The identification of early biomarkers of hyperinflammation may address the unpredictable onset of MIS-C, a feature shared by the development of Long Covid, another incompletely understood syndrome that occurs after SARS-CoV-2 infection.

Long Covid in children and young people

By **Binita Kane**[8] *and* **Sammie McFarland**[9]

Although most children recover fully from COVID-19, some will have persistent symptoms. Global prevalence estimates of Long Covid in children vary from 1:4 to 1:100 depending on cohort, methodology, and definition (*21*). The Office of National Statistics (ONS) estimates that ~120,000 children in the UK are suffering the effects of Long Covid, with 26,000 having symptoms for >1 year (*22*). A quarter of children experienced persistent symptoms months after hospitalization with acute COVID-19 infection, with almost 1 in 10 experiencing multisystem involvement (*23*). In total, over 200 different symptoms have been reported, affecting every organ system. Emerging evidence from case studies, patient-led charities, and pediatric services set up to treat Long Covid highlight that children are also affected by Long Covid after initial mild or asymptomatic illness (*24*). However, there are challenges in interpreting the existing literature, particularly with regards to defining "control groups" in a population that has had high levels of COVID-19 infection.

Owing to the heterogeneity of symptoms, diagnosing Long Covid in children is challenging, especially in the absence of established biomarkers of disease. The first research definition of Long Covid in children and young people was published recently (*25*). This acknowledges that the lack of complete recovery after acute infection with SARS-CoV-2 for >4 weeks is concerning, but symptoms persistent at 12 weeks are required to meet the definition of Long Covid. An international consensus for a pediatric Long Covid definition is anticipated in 2023; this will be an important step in the illness being recognized and managed consistently.

Children with Long Covid can be divided into those unable to return to normal life because of symptoms such as (but not limited to) pain, fatigue, post-exertional symptom exacerbation, headaches, and cognitive difficulties, and those who have additional postinfectious complications, such as pediatric multisystem inflammatory syndrome (PIMS), acute neurological disease, myocarditis, and other potential acute or subacute illnesses (*21*). Although the evidence for underlying biomedical causes of the long-term impact of SARS-CoV-2 infection in adults is growing quickly, there

Some children who were hospitalized with COVID-19, as well as some with asymptomatic or mild COVID-19, develop Long Covid.

PHOTO: HANNAH BEIER/REUTERS

is a paucity of similar studies in children, limiting our understanding of the disease in young people. Potential mechanisms underlying Long Covid include viral persistence in tissues, disorders of coagulation, immune dysfunction, and autoantibody production (*21*). At the population health level, there is a suggestion that the incidence of type 1 diabetes is increased in children with Long Covid, although the cause of this is unclear and requires further investigation (*26*). There are currently no longitudinal studies that report outcomes for recovery in children.

A substantial global biomedical research effort, alongside cooperation with patients and families, is required to understand the long-term impact of Long Covid in children and to develop effective services and therapies. Limiting severity of disease through vaccination and taking preventive steps to mitigate against airborne virus transmission, such as improving indoor ventilation in schools, remain important tools to prevent the burden of Long Covid increasing in children. This is especially important for low- and middle-income countries (LMICs) where additional factors such as poor sanitation, overcrowding, and malnutrition pose further risks to child health (*27*).

What is the role of children in transmission of SARS-CoV-2?

By **Deepti Gurdasani**[10] *and* **Christina Pagel**[11]

There is now unequivocal evidence that children play an important role in the transmission of SARS-CoV-2, particularly within school and household settings. The role that a person or group plays in transmission is dictated by a combination of exposure (how likely they are to get exposed to infection), susceptibility (how likely they are to get infected if exposed), and transmissibility (how likely they are to transmit to others if infected). Although the different contributions of these are difficult to disentangle (*28, 29*), the important role of children in SARS-CoV-2 transmission is likely due to higher levels of exposure in schools. Early studies probably underestimated susceptibility in children (*28–30*) by not considering both lower relative exposure in children during periods of lockdown and school closures, and that children are less likely to be symptomatic than adults (*31, 32*), and so are less likely to be identified as cases.

School exposure is high for two main reasons. Schools involve hundreds of children mixing daily in crowded indoor spaces that are often poorly ventilated, facilitating transmission. Additionally, infection in children is easily missed because they are more likely to have asymptomatic, relatively mild, or atypical symptoms compared to adults (*31, 32*); and they can be harder to test if COVID-19 is suspected. Studies have also shown that SARS-CoV-2 infection may go undetected among children, who can be "silent" asymptomatic spreaders in school outbreaks, which then spread into the community (*33*). Contemporaneous surveillance data from the UK ONS (based on random COVID-19 testing of households in England) (*32*) showed that symptom-based testing vastly underestimated actual case incidence and prevalence in children.

The clearest findings about the role of children in community transmission come from studies that showed substantial reductions and increases in pandemic growth when schools were closed and opened, respectively (*34*). Random survey data from the ONS (*32*) indicate repeated increases in, and spread from, school-age children into parental age groups, with increases in infection rates in children predating increases in other age groups after school opening. Very large studies of adults living with children

Children play an important role in SARS-CoV-2 transmission, and so mitigation measures at schools are especially important.

PHOTO: VNA XINHUA/EYEVINE/REDUX

in the US (*35*), UK (*36*), and Denmark (*37*) have shown a higher risk of infection among households with children. More recently, genomic studies have also confirmed superspreading events within schools that then spread infections back into the community (*33*). Fortunately, studies also show that robust multilayered mitigation measures within schools can greatly reduce school outbreaks and are associated with lower community prevalence (*35*).

To reduce the role that children play in transmission of SARS-CoV-2, and to limit the impact of COVID-19 on children's health and that of their families, it is important to reduce exposure and transmission through safer school environments (improving air quality through investing in better ventilation and air cleaning), using N95/FFP2 masks during high or increasing community transmission, vaccination, accessible case ascertainment (e.g., saliva testing), and public information on how to make homes safer environments. Key areas of future research include understanding reinfections in children (How often do they occur? What are the risks of Long Covid or severe disease?); variant-specific effects on immune escape, reinfection, and transmission; and developing vaccines or boosters that are safe for children but provide longer protection against infection, particularly for under-12s; quantifying the long-term benefits of reduction in airborne disease; and better air quality, through investment in cleaner air in schools.

The importance of vaccinating children against COVID-19

By **Peter Hotez**[12,13,14,15,16,17]

The rationale for vaccinating children is clear: The recent waves of COVID-19 caused by the Delta and Omicron variants have taught us that infections with SARS-CoV-2 are not as benign for children as previously thought. In the US, more than 1200 children have lost their lives due to COVID-19 since the pandemic began (*38*), a number equivalent to or higher than the 300 children who die annually from all vaccine-preventable illnesses (*39*). Globally, the United Nations Children's Fund (UNICEF) reports that ~17,000 children and adolescents have lost their lives to COVID-19, with ~50% of those being under the age of 10 years (*40*). However, this number is based on 4.4 million overall global COVID-19 deaths from 91 countries, which is generally considered an underestimate. The new World Health Organization determination of 14.9 million excess global deaths (*41*), together with a UNICEF finding that 0.4% of COVID-19 deaths occur among children and adolescents (*40*), could suggest that as many 60,000 children and adolescents have died so far. This number would place pediatric COVID-19 deaths on par with those caused by some other vaccine-preventable diseases such as pediatric meningitis and other illnesses caused by the bacterium *Haemophilus influenzae* type b (Hib) and almost as high as measles-related deaths (*42*).

COVID-19 vaccines are now approved or increasingly available for immunizing children over the age of 6 months. Beyond their potential to reduce deaths from COVID-19 are the actual measured benefits of vaccinating children. In the US Delta variant wave in the summer–fall of 2021, most hospitalizations occurred among unvaccinated adolescents—only 8% of hospitalized adolescents were fully vaccinated—although this number rose to just over 20% when Omicron infections first accelerated in the final month of 2021 (*43*). Another study found that two doses of the Pfizer-BioNTech mRNA COVID-19 vaccine were over 90% effective at protecting against MIS-C (*44*). Vaccinated children are also less susceptible to COVID-19 infection overall (*45*), although the current vaccines are less effective in children against symptomatic infection with the Omicron variant (*46*). Still missing are data on vaccinating children to prevent Long Covid or its neurological complications. However, based on evidence from adults, some pediatric societies currently advocate for pediatric COVID-19 vaccinations on this basis (*47*). Also lacking are studies on vaccine access, equity, and effectiveness for children with intellectual and developmental disabilities. Additionally, extensive pharmacovigilance studies in the US and Europe have so far confirmed the safety of pediatric COVID-19 immunizations, but such studies need to be extended to LMICs.

Despite the benefits of pediatric COVID-19 vaccinations and the enthusiastic support from most pediatric societies, parental acceptance for vaccinating their children remains low. For instance, in many US Southern states, only 14 to 15% of children aged 5 to 11 have received a COVID-19 vaccination. Adding to these concerns is rising antivaccine activism that targets all pediatric vaccinations, including those for COVID-19. Therefore, there is urgency to accelerate educational efforts for parents about the seriousness of COVID-19 infections in children and the positive health impact of immunizations. In parallel, expanded studies on vaccine effectiveness are needed for pediatric COVID-19 vaccinations. This is true for both the recently authorized mRNA vaccines and newer and next-generation platform technologies (including universal coronavirus vaccines). These vaccines might eventually be incorporated into routine childhood immunization programs to avoid future coronavirus infection vulnerabilities among new birth cohorts. Pediatric COVID-19 immunizations must be prioritized by policymakers and practitioners committed to global child health.

PHOTO: GERARDO VIEYRA/NURPHOTO/GETTY IMAGES

Original antigenic sin and childhood immune responses against SARS-CoV-2

By **Sarah Cobey**[18] *and* **Scott E. Hensley**[19]

Children's first exposures to SARS-CoV-2, whether through infection or vaccination, might shape the specificity of their immune responses to SARS-CoV-2 variants for the rest of their lives. Previous studies on influenza virus infections may inform how childhood SARS-CoV-2 infections affect long-lived immunity. Thomas Francis Jr. coined the phrase "original antigenic sin" in 1960 to describe the observation that antibody responses elicited by childhood influenza virus infections can be recalled later in life upon exposure to antigenically distinct influenza viruses (*48*). This recall can lead to different patterns of antibody specificity among individuals exposed to the same influenza virus strain. But it remains uncertain whether de novo immune responses to new antigens are diminished by these previous exposures and recalled immune memory. The specificity of antibody responses to influenza virus arises from targeting different epitopes. Often, these responses target epitopes that are conserved between contemporary and childhood viral strains. This bias can be beneficial when antibodies are neutralizing, but potentially detrimental when these epitopes change in the virus, leading to a loss of antibody recognition and greater risk of infection (*49*). Evidence suggests that variation in risk of medi-

Vaccination of children and adolescents can prevent death and severe disease and may also reduce the occurrence of Long Covid.

cally attended influenza virus infection and vaccine effectiveness between influenza seasons and age groups arises partly from these idiosyncratic interactions with immune memory (*50*–*54*).

Is original antigenic sin relevant to COVID-19 vaccination and SARS-CoV-2 infections of children? Although the answer is currently unknown, it is likely that the specificity of antibody responses to a SARS-CoV-2 variant is influenced by exposure history. For example, Omicron infection elicits a broader neutralizing antibody response in vaccinated adults compared to those who are unvaccinated, leading to higher neutralizing antibody titers to Omicron and Delta strains in vaccinated individuals (*55*). The specificity and magnitude of antibodies elicited by SARS-CoV-2 infection appear to be influenced by vaccination status (*56*), but it is difficult to interpret differences in immune data between vaccinated and unvaccinated individuals because unvaccinated individuals typically encounter more antigen during their infections. Furthermore, it is unknown how different exposures affect immune memory over the long term in children and adults.

To understand the role of antigenic sin in SARS-CoV-2 immunity, it is important to continue to investigate the magnitude, specificity, and functionality of B and T cell responses elicited by sequential SARS-CoV-2 exposures in animal models and humans. It will also be important to directly compare different types of SARS-CoV-2 vaccines in people with different exposure histories. For example, mRNA vaccines might promote better de novo responses than some other vaccine platforms in individuals with prior SARS-CoV-2 immunity because these vaccines elicit very long-lived germinal centers (where B cells evolve to become high-affinity plasma cells that produce specific antibodies) (*57*) that prolong the availability of antigens to both naïve and memory B cells. Longitudinal studies are important to elucidate the short- and long-term effects of different SARS-CoV-2 exposures on immune memory, infection risk, and vaccine effectiveness in children.

REFERENCES AND NOTES

1. H. Q. McLean *et al.*, *Pediatrics* **149**, e2021054178 (2022).
2. C. A. Pierce *et al.*, *Sci. Transl. Med.* **12**, eabd5487 (2020).
3. C. A. Pierce *et al.*, *JCI Insight* **6**, e148694 (2021).
4. S. Madera *et al.*, *Sci. Rep.* **11**, 3044 (2021).
5. S. P. Weisberg *et al.*, *Nat. Immunol.* **22**, 25 (2021).
6. M. Yoshida *et al.*, *Nature* **602**, 321 (2022).
7. J. Loske *et al.*, *Nat. Biotechnol.* **40**, 319 (2022).
8. L. C. Rowntree *et al.*, *Immunity* **55**, 1299 (2022).
9. L. R. Feldstein *et al.*, *JAMA* **325**, 1074 (2021).
10. L. R. Feldstein *et al.*, *N. Engl. J. Med.* **383**, 334 (2020).
11. T. Alsaied *et al.*, *Circulation* **143**, 78 (2021).
12. J. Chou, P. G. Thomas, A. G. Randolph, *Nat. Immunol.* **23**, 177 (2022).
13. M. Holm *et al.*, *JAMA Pediatr.* **176**, 821 (2022).
14. N. Levy *et al.*, *JAMA* **327**, 2452 (2022).
15. L. A. Vella *et al.*, *Sci. Immunol.* **6**, eabf7570 (2021).
16. C. R. Consiglio *et al.*, *Cell* **183**, 968 (2020).
17. V. Sancho-Shimizu *et al.*, *J. Exp. Med.* **218**, e20210446 (2021).
18. G. B. Sigal *et al.*, *Clin. Infect. Dis.* 10.1093/cid/ciac160 (2022).
19. M. B. Salzman, C.-W. Huang, C. M. O'Brien, R. D. Castillo, *Emerg. Infect. Dis.* **27**, 1944 (2021).
20. L. D. Zambrano *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **71**, 52 (2022).
21. D. Buonsenso *et al.*, *Future Microbiol.* **17**, 551 (2022).
22. Office for National Statistics (ONS), Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK (ONS, London, 2022).
23. I. M. Osmanov *et al.*, *Eur. Respir. J.* **59**, 2101341 (2022).
24. D. Buonsenso *et al.*, *Future Microbiol.* **17**, 577 (2022).
25. T. Stephenson *et al.*, *Arch. Dis. Child.* **107**, 674 (2022).
26. C. Kamrath *et al.*, *Diabetes Care* **45**, 1762 (2022).
27. H. J. Zar, J. Dawa, G. B. Fischer, J. A. Castro-Rodriguez, *Paediatr. Respir. Rev.* **35**, 70 (2020).
28. Z. Hyde, *Med. J. Aust.* **214**, 190 (2021).
29. Z. Hyde, *Clin. Infect. Dis.* **74**, 152 (2022).
30. Q. Bi *et al.*, *Lancet Infect. Dis.* **20**, 911 (2020).
31. P. Poletti *et al.*, *JAMA Netw. Open* **4**, e211085 (2021).
32. M. Hippich *et al.*, *Med* **2**, 149 (2021).
33. C. Meuris *et al.*, , *JAMA Netw. Open* **4**, e2128757 (2021).
34. N. Haug *et al.*, *Nat. Hum. Behav.* **4**, 1303 (2020).
35. J. Lessler *et al.*, *Science* **372**, 1092 (2021).
36. H. Forbes *et al.*, *BMJ* **372**, n628 (2021).
37. A. Husby, G. Corn, T. G. Krause, *MedRxiv* 02.28.21250921 (2021).
38. National Center for Health Statistics, Provisional COVID-19 Deaths: Focus on Ages 0-18 Years; https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-Focus-on-Ages-0-18-Yea/nr4s-juj3, accessed 2 July 2022.
39. Office of Disease Prevention and Health Promotion, Immunization and Infectious Diseases; https://www.healthypeople.gov/2020/topics-objectives/topic/immunization-and-infectious-diseases, accessed 2 July 2022.
40. UNICEF, COVID-19 confirmed cases and deaths, age- and sex-disaggregated data. March 2022; https://data.unicef.org/resources/covid-19-confirmed-cases-and-deaths-dashboard/, accessed 2 July 2022.
41. World Health Organization, 14.9 million excess deaths associated with the COVID-19 pandemic in 2020 and 2021, 5 May 2022; https://www.who.int/news/item/05-05-2022-14.9-million-excess-deaths-were-associated-with-the-covid-19-pandemic-in-2020-and-2021, accessed 2 July 2022.
42. L. D. Frenkel, *Allergy Asthma Proc.* **42**, 378 (2021).
43. K. J. Marks *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **71**, 271 (2022).
44. L. D. Zambrano *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **71**, 52 (2022).
45. M. Yigit *et al.*, *Pediatr. Infect. Dis. J.* 10.1097/INF.0000000000003625 (2022).
46. C. J. Cohen-Stavi *et al.*, *N. Engl. J. Med.* **387**, 227 (2022).
47. M. Stein *et al.*, *Vaccines* **10**, 81 (2022).
48. T. Francis, *Proc. Am. Philos. Soc.* **104**, 572 (1960).
49. S. Cobey, S. E. Hensley, *Curr. Opin. Virol.* **22**, 105 (2017).
50. B. Flannery *et al.*, *J. Infect. Dis.* **218**, 189 (2018).
51. K. M. Gostic *et al.*, *PLOS Pathog.* **15**, e1008109 (2019).
52. P. Arevalo, H. Q. McLean, E. A. Belongia, S. Cobey, *eLife* **9**, e50060 (2020).
53. D. M. Skowronski *et al.*, *J. Infect. Dis.* **216**, 1487 (2017).
54. K. M. Gostic, M. Ambrose, M. Worobey, J. O. Lloyd-Smith, *Science* **354**, 722 (2016).
55. K. Khan *et al.*, *Nature* **607**, 356 (2022).
56. C. J. Reynolds *et al.*, *Science*, eabq1841 (2022).
57. N. Pardi *et al.*, *J. Exp. Med.* **215**, 1571 (2018).

ACKNOWLEDGMENTS

B.C.H. and K.C.H. are funded by NIH NIAID R01 AI134367. C.A.P. was supported by the Medical Scientist Training Program of the Albert Einstein College of Medicine (T32GM007288) and the institutional training grant in HIV/AIDS pathogenesis (T32AI007501). J.C. is supported by the National Institute of Diabetes and Digestive and Kidney Diseases to study MIS-C. C.P. and B.K. are members of Independent SAGE (www.independentsage.org/). P.H. is a coinventor of a COVID-19 recombinant protein vaccine technology owned by Baylor College of Medicine (BCM) that was recently licensed by BCM nonexclusively and with no patent restrictions to several companies committed to advance vaccines for LMICs. S.C. and S.E.H. are funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no. 75N93021C00015.

10.1126/science.ade1675

MEDICINE

Understanding myalgic encephalomyelitis

Myalgic encephalomyelitis and Long Covid have overlapping presentation

By **Sonya Marshall-Gradisnik** *and* **Natalie Eaton-Fitch**

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a severe condition characterized by post-exertional neuroimmune exhaustion (PENE) accompanied by neurological, immunological, gastrointestinal (GI), and mitochondrial disturbances (*1*). The global prevalence of ME/CFS is ~1%, affecting 17 million to 24 million people (*2*). ME/CFS is heterogeneous not only in symptom presentation but also illness trajectories, which can be worsening, plateauing, improving, or relapsing-remitting. Approximately 25% of patients with ME/CFS are considered severe and are bound to their homes. Although the etiology of ME/CFS is elusive, a large proportion of patients (~60%) report post-infectious onset, such as after Epstein-Barr virus infection (*3*). The recent emergence of a chronic post-infectious condition, called Long Covid, overlaps considerably with ME/CFS in immunological, mitochondrial, and neurological dysfunctions (*4*). These similarities have resulted in increased interest and acceptance of ME/CFS as a disease and may stimulate research, the development of a diagnostic test, and pharmacotherapeutic interventions in ME/CFS that may be applied to Long Covid.

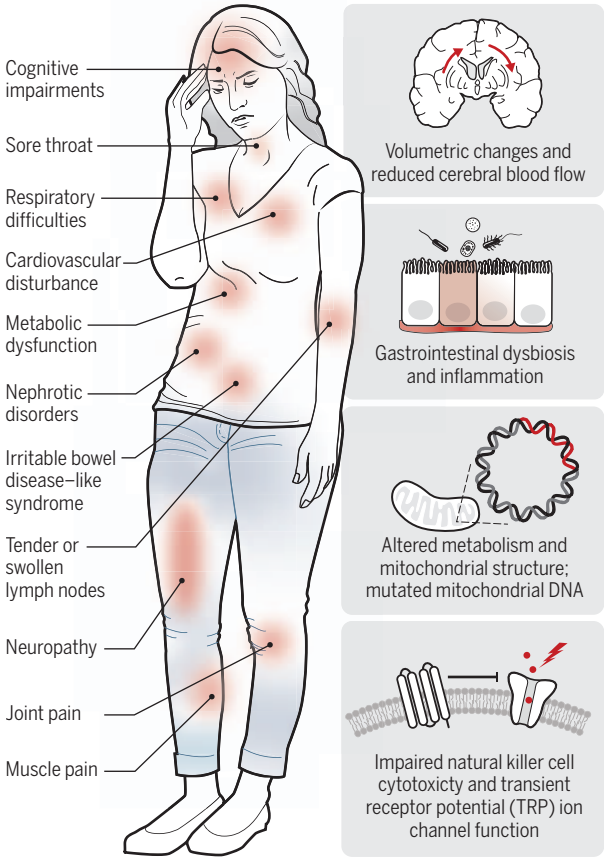
Neurological disturbances such as cognitive impairment, autonomic dysfunction, altered pain and sensory perception, and sleep disturbances are essential for diagnosis of ME/CFS and are commonly reported in Long Covid (see the figure). The World Health Organization (WHO) categorizes ME/CFS as a disease of the nervous system. Neuroimaging of ME/CFS patients has revealed anatomical, neurochemical, and functional brain changes. For example, brain magnetic resonance imaging (MRI) found global gray and white matter volume changes and also differences in the cortical and subcortical regional volumes in ME/CFS patients (*5*). Impaired brainstem connectivity identified with functional MRI (fMRI) and regression of white matter was associated with autonomic nervous system (ANS) measures in ME/CFS, including sleep disturbances and respiratory rate, which may lead to other symptoms, including pain, fatigue, impaired concentration and memory, and sensory and motor dysfunction (*6*). Such neurological symptoms are also commonly reported by Long Covid patients, and brain MRI has shown higher gray matter volumes in hippocampi that correlated with memory loss, a result also reported in ME/CFS (*4*). Overall, the underlying mechanism resulting in these MRI findings is unclear. Nonetheless, MRI and fMRI are important techniques to help elucidate the pathology of ME/CFS, as well as Long Covid.

Many ME/CFS and Long Covid patients report GI symptoms, including abdominal pain, nausea, constipation, and diarrhea associated with changes to motility (in which the activity of nerves and muscles of the GI tract are impaired) and dysbiosis (imbalanced gut bacteria) (*7*). Dysregulated gut-brain signaling is proposed to play a role in the neurological dysfunction observed in patients with ME/CFS. For example, production of neurotransmitters such as serotonin, dopamine, and γ-aminobutyric acid by enteroendocrine cells can be stimulated by intestinal bacteria through the production of neurotransmitter precursors, which may influence memory, cognition, mood, and sleep (*8*). However, this does not explain other symptoms, such as neurosensory and autonomic dysfunctions. Instead, the presence of GI dysregulation, through proposed disturbances in GI nerve and consequently smooth muscle activity, emphasizes the involvement of impaired ANS regulation in the symptomatology of ME/CFS. Yet, GI symptoms and irritable bowel are not specific to ME/CFS. An analysis reported that the role of dysbiosis and intestinal barrier permeability in ME/CFS and post-infectious syndromes remain unclear, and results are often unequivocal. Thus, further investigation using larger cohorts and consistent inclusion criteria is required; confounding variables such as irritable bowel disease should also be accounted for (*7*, *8*).

Cross talk between the nervous, GI, and immune systems has been reported in other post-infectious disease states, such as Long Covid. Pathological interactions between GI disturbances and the immune system may result from translocation of immunogenic bacteria into the bloodstream of ME/CFS patients (*9*). Immune dysfunction is a key feature of ME/CFS, and many symptoms indicate chronic immune activation—for example, persistent sore throat, tender and enlarged lymph nodes, and fever. The consistent findings of immune dysfunction, demonstrated

Proposed mechanisms underlying ME/CFS

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) presents as a range of symptoms that affect multiple organ systems. Enteric dysbiosis, neurological and immune dysfunction, as well as impaired mitochondrial function are implicated in the pathomechanism of ME/CFS. These symptoms also occur in Long Covid, although with differing prevalence.



GRAPHIC: N. CARY/*SCIENCE*

through aberrations in function and number of B- and T-lymphocytes and natural killer (NK) cells, suggest that ME/CFS is better classified as a neuroimmune disorder (*4*). It has also been suggested that ME/CFS is an autoimmune condition (*10*). A small number of studies in ME/CFS patients reported autoantibodies to neuroendocrine receptors and elevated $CD20^+$ and $CD21^+$ B lymphocytes, which are both commonly increased in patients with autoimmune diseases (*10*). No additional evidence of ME/CFS as an autoimmune disease has been reported. Furthermore, in a large ME/CFS cohort, clinical trials that used B-lymphocyte depletion found no improvement in symptoms (*11*). Changes in surface receptors, lytic proteins, degranulation, protein kinase phosphorylation, and cytokine production suggest that immune cell exhaustion plays a role in immunological symptoms of ME/CFS, including susceptibility to infection (*12*). Notably, T lymphocyte exhaustion also plays a role in recovery from COVID-19 and is associated with more severe outcomes for COVID-19 patients, and they persist in Long Covid. The only consistent immunological finding in ME/CFS patients is impaired NK cell cytotoxicity, which should be investigated further to understand the pathology of ME/CFS (*12*).

Immune dysfunction, cognitive and GI disturbances, and PENE in ME/CFS and Long Covid have been suggested to be associated, at least in part, with mitochondrial dysfunction. These include changes in mitochondrial structure, DNA, membrane potential, respiratory function, reactive oxygen species (ROS), antioxidant defense, metabolites, and coenzymes in plasma, urine, peripheral blood mononuclear cells (PBMCs), and isolated immune cells (*9*, *13*). Changes in metabolite concentrations suggest disturbances in cellular energy production (*9*, *13*). Conversely, mitochondrial adenosine triphosphate (ATP) concentrations and absolute ATP synthesis rates remained unchanged in PBMCs from ME/CFS patients compared with healthy controls (*9*). These findings are limited by the inclusion of patients who were diagnosed using criteria with low sensitivity for ME/CFS. Mitochondrial disturbances in both Long Covid and ME/CFS are believed to contribute to loss of cell activity. For example, lymphocytes exhibit impaired oxidative phosphorylation and increased ROS, contributing to immune exhaustion. However, further research is required to validate mitochondrial dysfunction as a biomarker of illness in patients with ME/CFS because current evidence is insufficient to determine causation.

National Centre for Neuroimmunology and Emerging Diseases, MHIQ, Griffith University, Gold Coast, Australia. Email: s.marshall-gradisnik@griffith.edu.au

An overlapping feature of these disruptions to the nervous, GI, and immune systems is calcium (Ca^{2+}) signaling. Ca^{2+} is a universal second messenger that affects numerous biological processes in all cell types (*9*). Transient receptor potential (TRP) channels are a superfamily of nonselective ion channels with high permeability to Ca^{2+}. Notably, TRP channels are widely expressed by cells composing multiple organ systems, including the nervous, musculoskeletal, cardiovascular, and GI systems, all of which are associated with symptoms reported by ME/CFS and Long Covid patients. The occurrence of ME/CFS as a result of abnormal ion channel function (a channelopathy) gained interest in the 1990s, and recently, TRP ion channels have been implicated in not only ME/CFS but also the pathology of Long Covid (*4*). Single-nucleotide polymorphisms have been identified to impair TRPM3 (also called melastatin) ion channel function and Ca^{2+} mobilization in NK cells of ME/CFS patients (*9*, *12*). Impaired TRP channel function and Ca^{2+} signaling is likely to impede cell and mitochondrial function, resulting in cognitive, immune, and GI manifestations. There is no animal model of ME/CFS; thus, NK cells provide an accessible cell model with which to investigate TRP pathology in vitro. However, alterations in TRP ion channels are yet to be validated as an underlying cause of ME/CFS.

The difficulty in identifying the mechanism of ME/CFS means that there are limited internationally accepted treatments for patients. Rintatolimod is an immunomodulatory drug that is used off-label to restore NK cell cytotoxicity in ME/CFS to improve symptoms associated with immune exhaustion, such as persistent sore throat and tender lymph nodes. However, because only mild improvements have been reported in ~40% of 412 patients across several clinical trials, rintatolimod has received orphan drug designation in the United States and is approved for severe cases of ME/CFS in limited countries (*11*). Other treatment interventions include antivirals, immunomodulators, analgesics, nutritional supplements, cognitive behavioral therapy, and graded exercise therapy to target specific symptoms. But a review reported insufficient evidence of the effectiveness of these interventions in improving symptoms (*11*), and many patients report susceptibility to side effects, thus posing further challenges to their care and studies of effective drugs (*14*). The lack of consistent findings is attributed to a low number of study samples, the use of nonspecific criteria for participation, and absent reporting of confounding variables, such as polypharmacy. More recently, promising results for the off-label use of low-dose naltrexone (LDN), an opioid receptor inhibitor used to treat drug and alcohol addiction, have been reported. Treatment of ME/CFS patients with LDN restored TRPM3 ion channel dysfunction in NK cells (*11*). Furthermore, 73.9% of 218 ME/CFS patients routinely taking LDN reported improved sleep and reduced pain and neurological disturbances (*15*). Although there is potential benefit in applying repurposed therapeutics in the treatment of ME/CFS, clinical trials are required to substantiate these claims.

An important research challenge is the lack of a validated biomarker, laboratory-based test, and animal model, likely attributed to inconsistency in protocols, such as cell isolation, sample type, and technique. These challenges provide important lessons for research of Long Covid. By developing interdisciplinary and consistent research protocols, the pathomechanism of ME/CFS and Long Covid can be elucidated. The recent interest in the overlap that exists between ME/CFS and Long Covid poses several questions, such as whether Long Covid predisposes a person to ME/CFS, and whether an individual has Long Covid or ME/CFS. Research into the etiology of ME/CFS and Long Covid should simultaneously identify the mechanism and diagnostic approach to both. ME/CFS poses a substantial health concern and has recently been taken more seriously since the emergence of Long Covid, and renewed focus on diagnostic, research, and treatment practices is needed. ■

REFERENCES AND NOTES

1. B. M. Carruthers *et al.*, *J. Intern. Med.* **270**, 327 (2011).
2. E.-J. Lim *et al.*, *J. Transl. Med.* **18**, 100 (2020).
3. T. L. Wong, D. J. Weitzer, *Medicina* (Kaunas) **57**, 418 (2021).
4. O. A. Sukocheva *et al.*, *J. Adv. Res.* 10.1016/j.jare.2021.11.013 (2021).
5. R. Maksoud *et al.*, *PLOS ONE* **15**, e0232475 (2020).
6. T. Nelson *et al.*, *Front. Neurol.* **12**, 769511 (2021).
7. A. Varesi, U.-S. Deumer, S. Ananth, G. Ricevuti, *J. Clin. Med.* **10**, 5077 (2021).
8. R. S. König *et al.*, *Front. Immunol.* **12**, 628741 (2022).
9. D. Missailidis, S. J. Annesley, P. R. Fisher, *Diagnostics* (Basel) **9**, 80 (2019).
10. F. K. F. Mensah, A. S. Bansal, B. Ford, G. Cambridge, *Neurophysiol. Clin.* **47**, 131 (2017).
11. P. L. Toogood, D. J. Clauw, S. Phadke, D. Hoffman, *Pharmacol. Res.* **165**, 105465 (2021).
12. N. Eaton-Fitch *et al.*, *Syst. Rev.* **8**, 279 (2019).
13. B. D. Paul, M. D. Lemle, A. L. Komaroff, S. H. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024358118 (2021).
14. National Institute for Health and Care Excellence (NICE), "Myalgic encephalomyelitis (or encephalopathy)/chronic fatigue syndrome: diagnosis and management. Guidance n.d.:87" (NICE, 2021).
15. O. Polo, P. Pesonen, E. Tuominen, *Fatigue* **7**, 207 (2019).

ACKNOWLEDGMENTS

The authors are funded by the National Health and Medical Research Council, the Stafford Fox Medical Research Foundation, McCusker Charitable Foundation, and the Buxton Foundation. The authors disclose Patent Cooperation Treaty (PCT) patent application no. WO2016176726A1 and PCT provisional application no. 2022902253.

10.1126/science.abo1261

METAMATERIAL

Chiral light sources get a helping hand

Resonant metasurfaces pave the way for more compact sources of pure chiral light

By **Andrew Forbes**

Chirality describes the inherent "handedness" of many physical systems—from the left-handed amino acids that form the building blocks of life to the right-handed spiral staircases in the medieval castles of Europe. Light, too, can be chiral, manifesting as right-circular or left-circular polarization states, in which the electric field rotates clockwise or anticlockwise in time and traces out a spiral trajectory as it propagates through space. Despite the impressive technological progress since the discovery of light's chirality hundreds of years ago, the practical challenge of producing chiral light from a compact source remains elusive. This is because the chiral interaction of light and matter is typically very weak and thus requires the use of bulky materials. On page 1215 of this issue, Zhang *et al.* (*1*) demonstrate a metasurface that acts as a tiny and controllable source of pure and directional chiral light.

Around the turn of the 20th century, English physicist John Henry Poynting correctly argued that circularly polarized light must carry spin angular momentum (SAM) analogous to the angular momentum of spinning tops. This SAM was later experimentally confirmed by American physicist Richard Beth. The subsequent development of atomic physics revealed that light can also carry orbital angular momentum (OAM), which is linked to the spatial structure of light (*2*) and characterized by twisted helical wavefronts and spiraling energy flows (*3*). This OAM of light, same as the SAM, has a handedness based on a clockwise or anticlockwise twist, which gives positive or negative OAM to each photon.

These two forms of angular momenta in light, SAM and OAM, are often found together and add up to the total angular momentum the light carries, although the contribution of each to the total cannot be easily discerned. To complicate things further, the "chiral" SAM and the "twisted" OAM also interact with each other through a process known as spin-orbit coupling, which occurs in even the simplest of optical systems (*4*). Luckily, this process can also be harnessed to control the handedness of one component by manipulating the other.

Many potential applications can benefit from the ability to control the handedness of chiral and twisted light. For example, in optical communication, OAM could offer a new dimension for encoding information. However, unlike the evident chiral asymmetry found in amino acids and castle staircases, light has no handedness preference. One can exploit excited atoms to produce light with a twist, either directly for OAM light or by modulating atoms by use of external magnetic fields, triggering SAM chiral light emission along the direction of the field. However, these approaches tend to require inconveniently intricate, inefficient, and bulky device designs.

For many applications, a light source must have compact integrated photonic solutions that are small. To this end, microsize OAM lasers have achieved handedness control with the help of a chiral pump light source (*5*), the use of symmetry in microring lasers (*6*), and chiral-controlled nanoscale metasurfaces (*7*). However, because of the weak chiral interaction of light and matter, existing designs for producing chiral (SAM) light do not lend themselves to such miniature solutions.

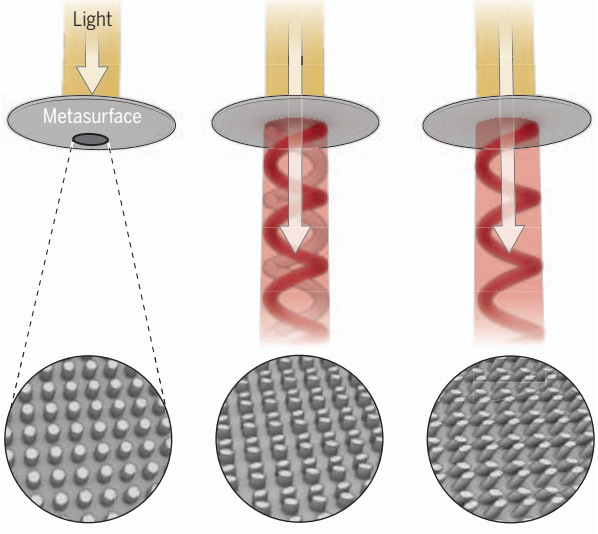
One promising approach to overcome the weak chiral light-matter interaction is to create artificial materials made from so-called "meta-atoms." These are not atoms in the usual sense. Instead, they are nanostructures that can be used to build desired three-dimensional (3D) metamaterials or 2D metasurfaces. These meta-atoms for enhancing chiral response have seen tremendous progress of late (*8*), but ultracompact solutions that offer high-purity chiral light have remained elusive (*9*), particularly for laser applications (*10*).

Zhang *et al.* tackle the problem with what at first seems a paradoxical approach. Their design uses an electromagnetic state that lies within a so-called bound state in the continuum (BIC). The BIC refers to a state in which light of a certain frequency is trapped bouncing around, such as inside a carefully designed structure of nanopillars with a high level of symmetry (*9*). Such BICs come with many benefits, including enhanced amplification and chiral filtering. However, to be useful as a chiral source, the trapped light must be freed. To achieve this, one can break the necessary symmetry for BICs at some desired frequency, which will make the light only partly bound.

The trade-off between the benefits of radiation (letting out the light) and the ability to manipulate the light property is a fine balancing act. This trade-off can be controlled by changing the nanopillar arrays. Previous research efforts have tried changing the orientation angle of the individual oval-shaped pillars by rotating them as viewed from the top (see the figure). However, the common

A high-purity miniature chiral light source

With potential applications in communications and sensing, a compact chiral light source with high purity in handedness has been elusive. Zhang *et al*. improved on existing design for a resonant metasurface that acts as a miniature chiral light laser.



When nanopillars on the metasurface are **perfectly round and upright**, they trap the light, making it useless as a light source.

Changing the shape and orientation of the nanopillars breaks their symmetry, which allows the metasurface to release the chiral light.

Zhang *et al.* further break the symmetry by **tilting the pillars**. The design produces chiral light with much higher selectivity.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

School of Physics, University of the Witwatersrand, Johannesburg, South Africa. Email: andrew.forbes@wits.ac.za

problem with this approach is that it results in incoherent broadband emissions with low purity (*10*).

Zhang *et al.* solve this by finding an additional way to change the nanopillars for manipulating the light. They tilt the oval-shaped pillars so that they are all at an angle. The combined symmetry breaking of the rotation and forward tilt of the pillars brings the chiral filtering action of the BIC to coincide with the frequency at which the partly bound BIC radiates. They show a way to calculate the optimal tilt angle for chiral light with >99% purity in handedness. Moreover, because the nanopillar metasurface is resonant to certain frequencies of light, it also acts like an amplifier for producing a laser. And because lasers prefer to output the dominant optical mode—the one with either the highest gain or the lowest loss—the resulting competition of the left- and right-handed states further enhances the chiral purity as the more dominant handedness suppresses the other.

The ability to control and harness chiral light in compact micro- and nanosize structures is still very much in its infancy. The device by Zhang *et al.* marks the first demonstration of a high-purity chiral microlaser and is the first to offer controlled handedness through its metasurface design. However, the emerging ability to tailor the functionality of metasurfaces in a rewritable fashion (*11*) may allow optically switchable chirality directly from on-chip sources. The powerful demonstration of chiral control by Zhang *et al.* will surely motivate further studies toward extending this to other degrees of freedom of light, with the promise of full on-chip control of structured light by structured matter. This is an exciting prospect for future optoelectronics and optomagnetic devices, where control beyond the intensity of light is gaining traction (*12*, *13*), which is crucial to moving photonics closer in functionality to that of electronics. ■

REFERENCES AND NOTES

1. X. Zhang *et al.*, *Science* **377**, 1215 (2022).
2. A. Forbes, M. de Oliveira, M. Dennis, *Nat. Photonics* **15**, 253 (2021).
3. S. Franke-Arnold, *Nat. Rev. Phys.* **4**, 361 (2022).
4. K. Y. Bliokh, F. J. Rodríguez-Fortuño, F. Nori, A. V. Zayats, *Nat. Photonics* **9**, 796 (2015).
5. N. Carlon Zambon *et al.*, *Nat. Photonics* **13**, 283 (2019).
6. Z. Zhang *et al.*, *Science* **368**, 760 (2020).
7. H. Sroor *et al.*, *Nat. Photonics* **14**, 498 (2020).
8. M. Hentschel, M. Schäferling, X. Duan, H. Giessen, N. Liu, *Sci. Adv.* **3**, e1602735 (2017).
9. D. C. Marinica, A. G. Borisov, S. V. Shabanov, *Phys. Rev. Lett.* **100**, 183902 (2008).
10. Y. Chen *et al.*, *Nat. Rev. Phys.* **4**, 113 (2022).
11. S. Abdollahramezani *et al.*, *Nat. Commun.* **13**, 1696 (2022).
12. Z. Ji *et al.*, *Science* **368**, 763 (2020).
13. K. Jana *et al.*, *Nat. Photonics* **15**, 622 (2021).

10.1126/science.add5065

CLIMATE CHANGE

El Niño's role in changing fauna

Past faunas show how El Niño frequency affects eastern Pacific community structure

By **Daniel H. Sandweiss**[1,2] *and* **Kirk A. Maasch**[1,3]

El Niño–Southern Oscillation (ENSO) is a major driver of interannual climate variability in the Pacific basin and plays a pivotal role in global climate (*1*). Among the many ENSO phenomena are the eastern Pacific events—commonly known as El Niño, which is synonymous with "baby Jesus" in Spanish, a name that reflects its occurrence around Christmastime. Because of the far-reaching consequences of these global climate phenomena, it is important to understand and predict the frequency and magnitude of El Niño events and their impact on resources. On page 1202 of this issue, Broughton *et al.* (*2*) provide an evaluation of El Niño's effects on vertebrate fauna over 12,000 years and characterize the intensity of human exploitation of these resources.

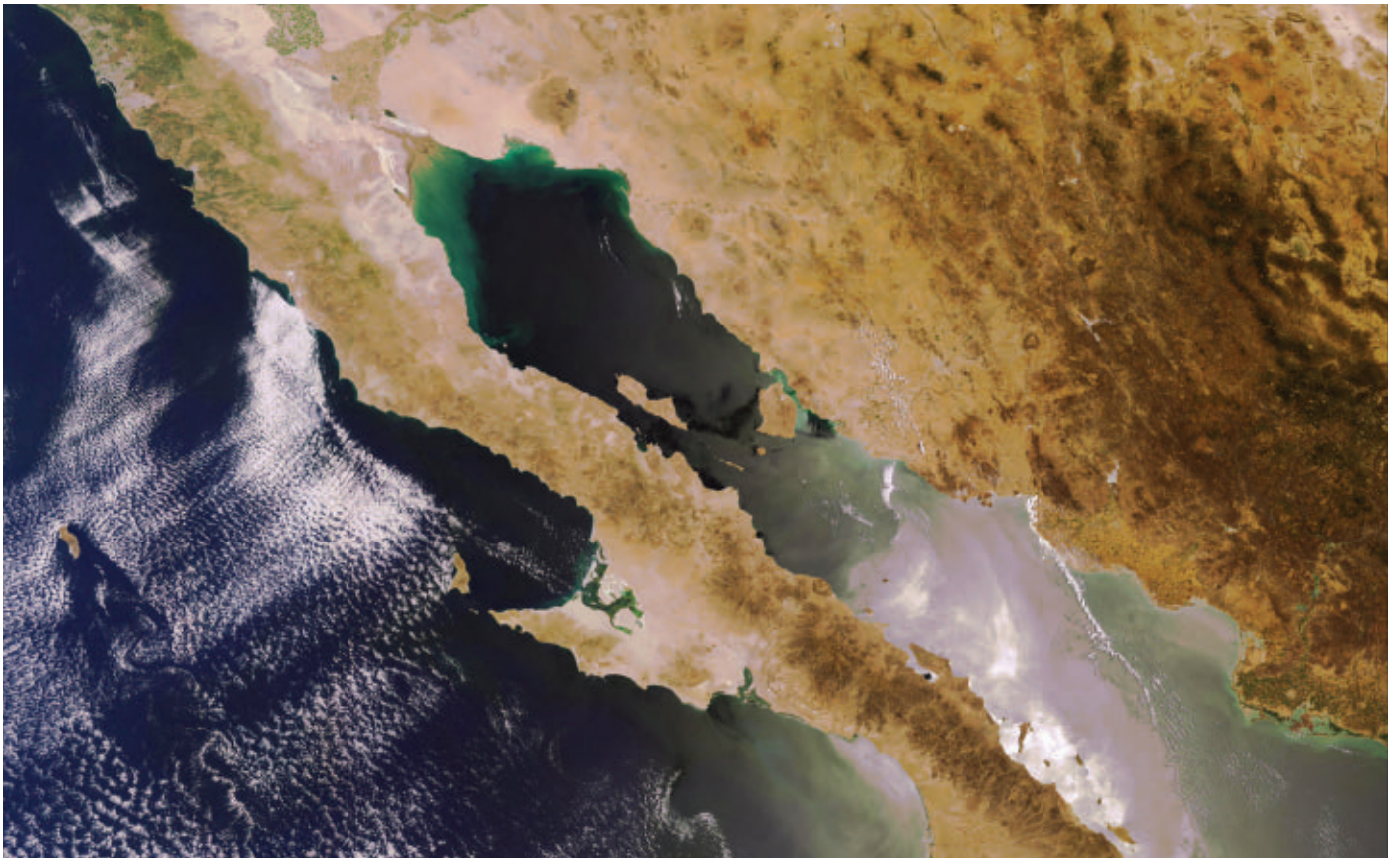
Climate modelers mostly predict an increase in El Niño events under global warming (*3*). However, these models need to be tested against data from the past, and written documentation only goes back to the mid-1800s. Because of this, the validation of long-term climate models often requires paleo records, many of which come from natural archives, including ice cores, lake cores, ocean cores, and corals. Archaeological records also offer insights into past climate at the local scale, often in areas and/or time periods that lack high-resolution natural archives. Broughton *et al.* reverse this approach by analyzing faunal remains from Abrigo de los Escorpiones in Baja California, Mexico, in comparison to a proxy record of past El Niño events. The authors use artifact counts as a proxy for human activity.

The site at Escorpiones offers the longest and best-resolved terrestrial record available to relate faunal change with fluctuating El Niño frequency, here derived from a record of sedimentation in Lake Pallcacocha in Ecuador (*4*). Broughton *et al.*'s analysis of their data revealed several important observations. They found that when El Niño was infrequent, particularly from 5 thousand to 7 thousand years (ka) ago, the fauna was highly variable and human activity was high. When El Niño reached a frequency above five events per century, vertebrate fauna was relatively stable, with low marine and high terrestrial productivity and decreased human activity. This makes sense because El Niño decreases nutrient upwelling off the coast of Baja California (*5*), which reduces marine biomass and increases precipitation (*6*). When marine resources become less abundant, human predation can be expected to have a greater effect on the scarcer marine fauna. Other studies have established ways to recognize overpredation in archaeological specimens, such as the observed effects of human predation on mollusk growth (*7*). Future studies can further explore these relationships, such as those between overpredation and declining productivity.

There have been similar studies based on archaeological records from the coast of Peru, which shares similar geographical features with Baja California. Both are arid deserts fronting a highly productive marine ecosystem that is supported by intense nutrient upwelling. Similar to its effects in Baja California, El Niño also reduces upwelling and brings torrential rainfall to Peru (*8*). The first indicators of past variability in El Niño frequencies came from the biogeography of archaeological mollusks, which was later supplemented with fish biogeography and stable-isotope analyses of mollusks (*9*) and fish otoliths (*10*). At Escorpiones, Broughton *et al.*'s results from molluscan biogeography and isotopes support the same historic frequencies of the El Niño record, but more such work would add nuance to that record.

The effects of El Niño frequency on marine fauna are not limited to the Escorpiones site and have been recognized elsewhere. For instance, a previous study found that in the 1900s, the average temperature of the Pacific was driven by El Niño frequency and correlated with a 50-year cycle in the relative abundance of sardines and anchovies, two species of tremendous economic importance (*11*). In coastal Peru, the reorganization of both molluscan and fish faunas accompanied

[1]Climate Change Institute, University of Maine, Orono, ME 04469, USA. [2]Department of Anthropology, University of Maine, Orono, ME 04469, USA. [3]Department of Earth and Climate Sciences, University of Maine, Orono, ME 04469, USA. Email: dan.sandweiss@maine.edu



Broughton *et al*. show how El Niño has affected marine biomass output and human activities over the past 12,000 years in Northern Baja California.

major transitions in El Niño frequency at 5.8 and 2.9 ka ago and led to the recognition of these frequency changes using archaeological records (*8*).

The results from Broughton *et al*. have clear implications for understanding the faunal response to El Niño frequency, which is likely to increase in the coming decades because of global warming (*1*). The Escorpiones faunal record also reflects well-known climate transitions in the past that placed Escorpiones in a broader context. One of the strongest climate signals over the past 12,000 years occurred at about 5 to 6 ka ago, when global climate change was characterized by polar cooling, tropical aridity, and major atmospheric circulation changes (*12*). In the context of global change, El Niño returned at low frequency to the Peruvian coast around 6 ka ago after a multimillennial hiatus or slowdown (*8*). This occurred in the middle of the transition in the Lake Pallcacocha record in Ecuador 5 to 7 ka ago and in the faunal and artifact records from Escorpiones in Baja California. A few thousand years later, around 2.9 ka ago, faunal remains and an ephemeral lake record from northern Peru suggest that El Niño frequency increased once again (*8*). In sync with this timeline, Broughton *et al*. found that a marine bird, *Chendytes lawi*, went locally extinct around 2.5 ka ago, likely because of the synergistic effects of predation and stress from increased El Niño frequency during that time.

The findings of Broughton *et al*. point to important research avenues to increase understanding of El Niño effects on resources of vital importance to humans. It is critical to recover more such long, continuous, well-dated records. In Peru, which shares many El Niño–related environmental characteristics with Baja California, there is no known site with an equivalently long, continuous, and well-resolved record of El Niño's impact on faunal resources. Future analyses should also include more invertebrate and vertebrate species as well as isotopic and more-standard zooarchaeological criteria (*13*).

The economic impact of large magnitude El Niño events is devastating in many parts of the world, although other regions may experience short-term benefits (*14*). The latest large-magnitude global event took place in 2015–2016 and brought a record-smashing hurricane season in the central north Pacific, water rationing in Puerto Rico, the biggest 1-year jump in atmospheric carbon dioxide concentrations on record, severe drought in Ethiopia, and the hottest global surface temperature on record up to that time (*15*). El Niño is sometimes called "the naughty child" because of the climate-driven disasters it often brings. If the past is the key to the future, studies such as that of Broughton *et al*. offer tools for better predicting what this naughty child may do in the coming centuries. ■

REFERENCES AND NOTES

1. V. Mendi, G. Dwarakish, *Int. J. Earth Sci. Eng* **8**, 130 (2015).
2. J. M. Broughton *et al*., *Science* **377**, 1202 (2022).
3. M. J. McPhaden, A. Santoso, W. Cai, Eds., *El Niño Southern Oscillation in a Changing Climate* (Geophysical Monograph Series 253, Wiley, ed. 1, 2021).
4. C. M. Moy, G. O. Seltzer, D. T. Rodbell, D. M. Anderson, *Nature* **420**, 162 (2002).
5. R. K. Takesue *et al*., *J. Geophys. Res*. **109**, C03009 (2004).
6. R. A. Minnich, E. F. Vizcaino, R. J. Dezzani, *Atmosfera* **13**, 1 (2000).
7. J. C. Castilla, *Trends Ecol. Evol*. **14**, 280 (1999).
8. D. H. Sandweiss *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **117**, 8271 (2020).
9. M. Carré *et al*., *Science* **345**, 1045 (2014).
10. C. F. T. Andrus, D. E. Crowe, D. H. Sandweiss, E. J. Reitz, C. S. Romanek, *Science* **295**, 1508 (2002).
11. F. P. Chavez, J. Ryan, S. E. Lluch-Cota, M. Ñiquen C., *Science* **299**, 217 (2003).
12. P. A. Mayewski *et al*., *Quat. Res*. **62**, 243 (2004).
13. E. J. Reitz, E. S. Wing, *Zooarchaeology* (Cambridge Univ. Press, ed. 2, 2008).
14. P. Cashin, K. Mohaddes, M. Raissi, *Finance Dev*. **53**, 30 (2016).
15. R. Lindsey, 2015 State of the Climate: El Niño came, saw, and conquered (2016); www.climate.gov/news-features/understanding-climate/2015-state-climate-el-niño-came-saw-and-conquered.

10.1126/science.add8890

PHOTO: ESA/CC BY-SA 3.0 IGO

EVOLUTION

Scaling brain neurogenesis across evolution

A genetic change could explain increased cortical neurogenesis in modern humans

By **Brigitte Malgrange** *and* **Laurent Nguyen**

The neocortex, the outer region of the cerebral cortex, is an evolutionarily advanced brain structure that is responsible for cognitive abilities. It has expanded in size and function across the mammalian clade (*1*). The extraordinary cognitive abilities of humans are thought to rely on brain size (and thus the number of neurons) and the intricate cytoarchitecture of the neocortex. The expansion and folding of the neocortex have been partly attributed to the existence of basal radial glial cells (bRGs). These progenitors generate most cortical neurons, and their number increases in gyrencephalic mammals (which have neocortical folds), such as primates and ferrets. On page 1170 of this issue, Pinson *et al.* (*2*) report that expression of a variant of human transketolase-like protein 1 (TKTL1) increases the number of bRGs in modern humans and thereby the output of upper layer projection neurons. This genetic change could contribute to differences in cognition with extinct archaic humans.

Among the hominids, humans have a neocortex that is larger than the neocortex of chimpanzees and bonobos and contains around twice the number of neurons (*3*). The emergence of human-specific genes has contributed to the expansion and rapid evolution of the cerebral cortex (*4*). Recent anthropological data obtained from skull endocasts (i.e., casting the skull cavity) suggest that Neanderthals share comparable brain volume with modern humans (*5*). Whether this reflects an equivalent amount of white (i.e., glial cells) and gray (i.e., neurons) matter and thus a similar production of the corresponding number of cortical neurons remains unknown.

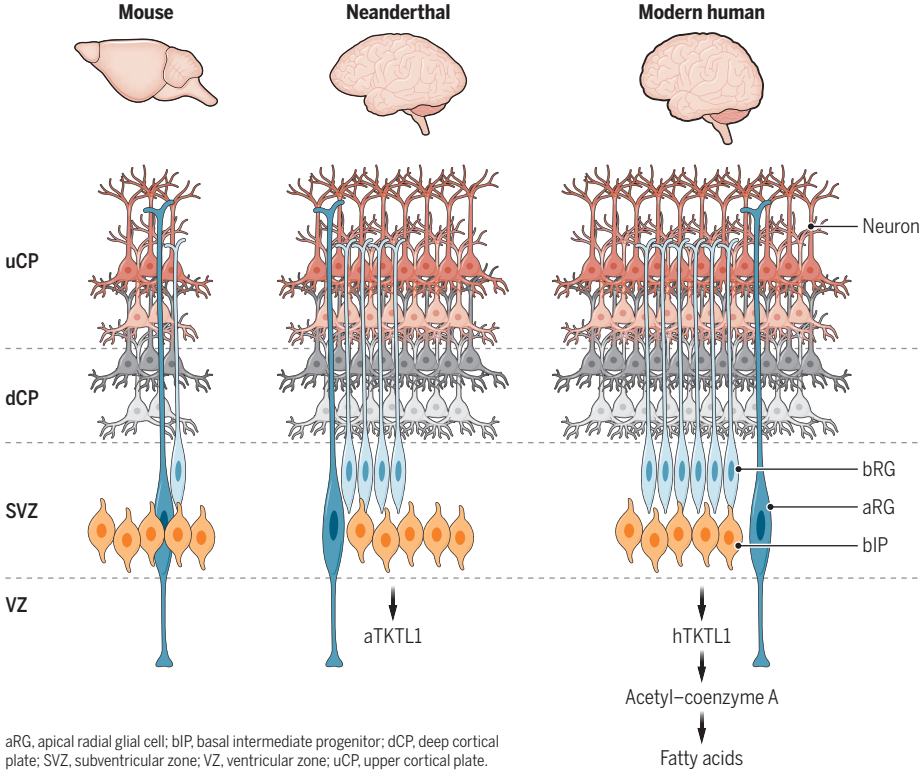
Because human brain fossil records are rare, efforts to understand the evolution of the neocortex at a cellular and molecular level have been limited to comparing living species (*6*). Observations of the developing neocortex in humans, nonhuman primates, and carnivores reveal how differences in neural progenitor cell populations can result in the variable size and shape of neocortices. Most current knowledge about the cellular and molecular mechanisms of neocortical development is based on experimental analysis of mouse models, whose neocortex exhibits key features that are generalizable to mammals, including a six-layered organization and separation into specialized regions. However, mice are not gyrencephalates, and their small size limits their relevance to human corticogenesis. Increases in neocortical surface and brain volume in gyrencephalates result from expanding progenitor cells in the outer subventricular zone during development (*7*–*9*). To what extent this cellular mechanism has been progressively established during the evolution of the genus *Homo* remains unknown. Pinson *et al.* elegantly explored this question by combining analyses of human cerebral organoid cultures with in vivo studies of mouse, ferret, and human brains.

What makes the modern human brain distinctive? Pinson *et al.* compared genomic sequences from modern with those of archaic humans, the Neanderthals. They found that an amino acid substitution encoded in the *TKTL1* gene of modern humans (*hTKTL1*) compared with the corresponding archaic human sequence (*aTKTL1*). *TKTL1* encodes a key enzyme in the nonoxidative branch of the pentose phosphate pathway (PPP), an essential metabolic pathway for acetyl–coenzyme A (acetyl-CoA) generation, which is used in fatty acid synthesis. TKTL1 has mainly been studied for its contribution to tumor metabolism because it acts as an enhancer of aerobic glycolysis (*10*). Notably, a *TKTL1* variant has recently been associated with neurodevelopmental disorders (*11*).

To decipher the role of TKTL1 during embryonic neurogenesis, Pinson *et al.* took

A genetic change increases human corticogenesis

A single Lys-to-Arg substitution in transketolase-like protein 1 (TKTL1) in modern humans (hTKTL1) compared with Neanderthal (aTKTL1) alters the metabolism of basal radial glial cells (bRGs). Expression of the hTKTL1 variant increases the number of bRGs and the upper cortical layer neuronal population. This is thought to contribute to the increased cognition of modern humans.



aRG, apical radial glial cell; bIP, basal intermediate progenitor; dCP, deep cortical plate; SVZ, subventricular zone; VZ, ventricular zone; uCP, upper cortical plate.

GRAPHIC: A. MASTIN/*SCIENCE*

GIGA-Stem Cells and GIGA-Neurosciences, Interdisciplinary Cluster for Applied Genoproteomics (GIGA-R), University of Liège, Liège, Belgium.
Email: bmalgrange@uliege.be; lnguyen@uliege.be

advantage of the mouse brain, which lacks expression of TKTL1, and overexpressed hTKTL1 or aTKTL1 in the apical progenitors that give rise to bRGs and neurons. They demonstrated that only hTKTL1 selectively promotes the proliferation and expansion of bRGs in the mouse embryonic brain. *hTKTL1* messenger RNA (mRNA) is also abundantly expressed in apical radial glia, but its overexpression does not affect this cell population. Moving to the gyrencephalic ferret neocortex, which endogenously expresses aTKTL1, Pinson *et al.* also showed that the bRG population was increased when hTKTL1 was overexpressed. Moreover, disrupting hTKTL1 expression in fetal human neocortical tissue in vitro reduced bRG abundance, whereas gene-edited human embryonic stem cells expressing aTKTL1 exhibit reduced bRG and neuron generation compared with those expressing hTKTL1 (see the figure).

What makes hTKTL1 functionally distinct from aTKTL1? Pinson *et al.* showed that hTKTL1 acts through the PPP. Whether aTKTL1 plays the same role has not been determined. The authors showed that overexpression of hTKTL1 but not aTKTL1 significantly increased acetyl-CoA amounts in bRGs. Together, the results support the hypothesis that hTKTL1 promotes the proliferation of bRGs, a cell population characterized by long basal processes that especially require high amounts of fatty acids. This mechanism could represent one crucial driver of brain evolution between ancestral and modern humans. Recently, an archaic amino acid substitution in the RNA splicing factor neuro-oncological ventral antigen 1 (NOVA1) was found to substantially alter cell proliferation and synaptogenesis in human brain organoids (*12*). Together, these observations open the path to discovering more-specific evolutionary changes that shaped the modern human brain and may also help us predict the next steps of its evolution. ■

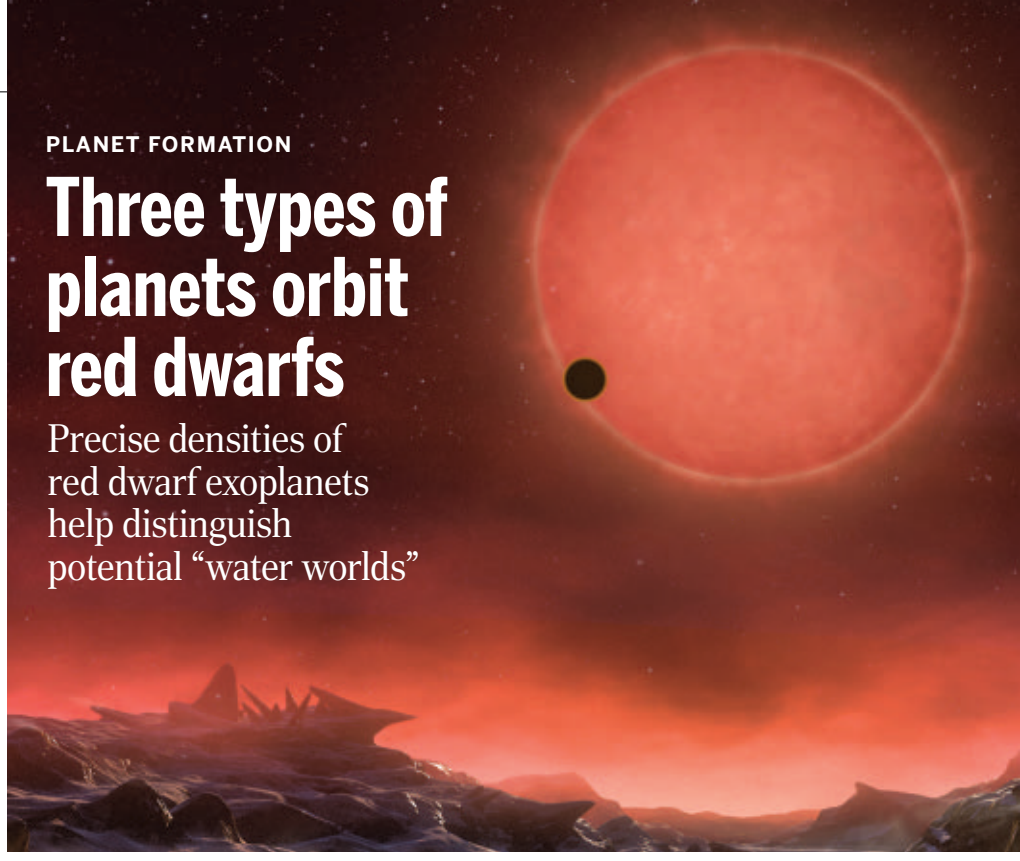
REFERENCES AND NOTES

1. A. Espinós, E. Fernández-Ortuño, E. Negri, V. Borrell, *Dev. Neurobiol.* **82**, 428 (2022).
2. A. Pinson *et al.*, *Science* **377**, eabl6422 (2022).
3. S. Herculano-Houzel, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10661 (2012).
4. M. Florio *et al.*, *Science* **347**, 1465 (2015).
5. C. VanSickle, Z. Cofran, D. Hunt, *Am. J. Phys. Anthropol.* **173**, 768 (2020).
6. S. J. Gould, *Science* **196**, 517 (1977).
7. I. H. Smart, C. Dehay, P. Giroud, M. Berland, H. Kennedy, *Cereb. Cortex* **12**, 37 (2002).
8. D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, *Nature* **464**, 554 (2010).
9. S. A. Fietz *et al.*, *Nat. Neurosci.* **13**, 690 (2010).
10. S. Langbein *et al.*, *Br. J. Cancer* **94**, 578 (2006).
11. A. Azzarà *et al.*, *Childs Nerv. Syst.* **37**, 2589 (2021).
12. C. A. Trujillo *et al.*, *Science* **371**, eaax2537 (2021).

ACKNOWLEDGMENTS

The authors are supported by the Fonds National de la Recherche Scientifique (FNRS) (WELBIO CR-2022A-12 to L.N. and PDR – 33666709 to B.M.).

10.1126/science.ade4388



PLANET FORMATION

Three types of planets orbit red dwarfs

Precise densities of red dwarf exoplanets help distinguish potential "water worlds"

ILLUSTRATION: ESO/M. KORNMESSER/CC BY 4.0

By **Johanna Teske**

Small planets orbiting faint red stars less than half the size of the Sun are numerous and could be the best places to search for signatures of life (*1*). However, the detailed compositions of these planets—and the implications for how they formed and evolved—have been difficult to study because of how far away they are from Earth and because of the red light that their stars emit. On page 1211 of this issue, Luque and Pallé (*2*) present refined compositions of small planets orbiting around red dwarf stars. They find evidence that the planets fall into three main types: rocky, watery (including icy), and gassy. This result differs from most previous studies of small planets that have suggested only rocky and gassy types (*3*, *4*). Although the presence of watery small exoplanets is particularly enticing, all three types of planets around red dwarfs could present potentially habitable conditions for life.

One of the most important recent findings in exoplanet science is the discovery of a bimodal distribution in the size of small planets around Sun-like stars (*3*). This bimodal distribution is thought to be characterized by planets with or without a hydrogen and/or helium atmosphere, with the two "modes" being gassy and rocky (*4*). Researchers could distinguish these two planet types on the basis of precise measurements of planetary radii and orbital periods using the Kepler space telescope and the Gaia space observatory. However, without knowing the masses of these planets, it is difficult to estimate their densities and thus know their compositions (rocky or gassy) with certainty. Mass measurements typically require high-resolution instruments on ground-based telescopes, and these measurements tend to be more constrained to exoplanets around brighter host stars. In all, only ~10% of detected small exoplanets have both radius and mass measurements.

Luque and Pallé took advantage of newly detected planets found around closer, and thus relatively brighter, red dwarf stars from the Transiting Exoplanet Survey Satellite (TESS). Their sample included 34 exoplanets with precise radius and mass measurements. Based on their sample, they found evidence for three different density regimes, which include so-called water worlds, with a 50/50 rock-to-ice ratio, in addition to the rocky and gassy planets. Their findings suggest that the water world planets tend to form with ice and rock far away from the host star before migrating to an orbit closer to the host star (see the figure).

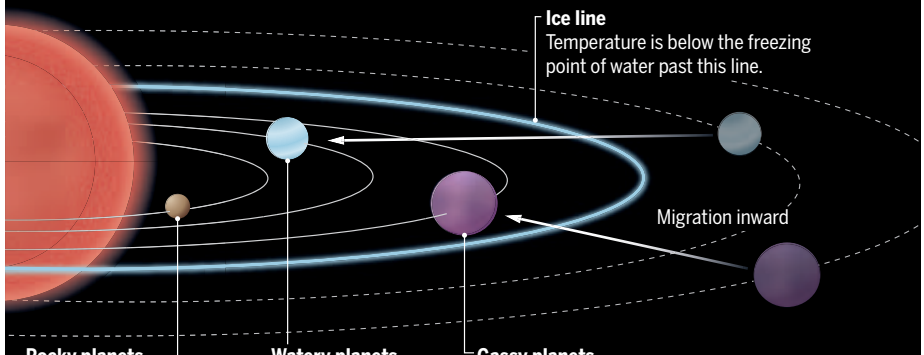
The discovery of watery planets in other planetary systems is itself not a big surprise because water is commonly detected around young stars (*5*). Within our Solar System, asteroids and comets left over from planet formation are water rich (*6*), and numerous moons and dwarf planets have shown signs of subsurface oceans and/or ice layers (*7*). Established planet formation models have indeed predicted that some planets

Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA.
Email: jteske@carnegiescience.edu

Rocky, watery and/or icy, and gassy planets can all form around a red dwarf star.

How small planets form around red dwarf stars

Planet formation differs among different solar systems. Luque and Pallé present the possible pathways for rocky, watery and/or icy, and gassy planets to form around red dwarf stars. Not all red dwarf stars have all three types of planets, but they are all shown here around one star for conciseness.



Rocky planets
A small rocky planet can form from the relatively dry material close to the host star. These planets have densities that are mostly consistent with an Earth-like density.

Watery planets
A watery and/or icy planet with a 50/50 rock-to-ice ratio can form outside the ice line and migrate inward toward the host star.

Gassy planets
A gassy planet may form similarly to a watery and/or icy planet, but it has more mass initially and thus is able to gather hydrogen and helium around it before migrating inward. However, the nature and formation of gassy planets are not well settled, and there are multiple competing models.

*Stars and planets not to scale

GRAPHIC: N. CARY/*SCIENCE*

tend to form beyond the ice line—that is, at a distance far enough from the host star where water exists as ice—before migrating closer to the star. With this in mind, it may be particularly easy for water-rich planets to form around red dwarf stars because those planetary systems seem to be more efficient at forming planets at far distances and migrating them inward before they gather too much hydrogen and/or helium gas from their surroundings and become gassy (*8*).

Even if a small planet does form within the ice line of a red dwarf system, icy materials can still crash into the planet later on, because these systems are less likely to have giant planets blocking the inward migration of icy materials (*9*). Based on this assumption, planet formation models predict that small inner planets with one or more giant outer planets in the same planetary system should be relatively dry (*10*). This prediction is in contrast with the likely origin of Earth's water, which involves Jupiter deflecting icy materials from our outer Solar System toward Earth (*11*). These competing perspectives will need to be compared with observations to reveal what dictates the water budgets of small planets around other stars and whether our Solar System is an exception to the rule.

There is still debate as to whether small planets that orbit close to the host star all form by a common pathway, with rocky cores and an atmosphere that sometimes is lost because of heating from the star and/or inside the planet itself. This debate is informed by the relationship between planet radius (R_p) and orbital period (P), where the strength of the linear correlation of this relationship is a pivotal factor in formation models. For red dwarf planets, the results have been mixed, with claims for a different (*12*) or similar (*13*) formation process. Although the R_p-P trend found in the red dwarf planets presented by Luque and Pallé is also inconclusive, the authors argue that planet density, which considers mass in addition to size, may be a more indicative metric.

Planet density was used in a recent more-complex and rigorous study that tested how well different mixtures of planet types match existing data (*14*). This other study modeled planet occurrence rates as a function of radius, mass, and period and found support for models of the observed planet population both with and without water worlds. However, the study only considered planets around Sun-like stars and not red dwarfs. The existing data on Sun-like systems are promising for robustly distinguishing planet types, but improved data and modeling approaches are needed to fully leverage the additional information from small planet densities, especially around red dwarf stars.

According to Luque and Pallé, red dwarf systems are capable of forming rocky, watery, and gassy planets, and at least some of the watery planets are formed past the ice line before migrating inward. There will be further evidence to test this finding as more data become available in the near future. Several of the water world planets identified by Luque and Pallé already have planned observations that will inform researchers of the presence and compositions of their atmospheres. In the coming years, more nearby, small red dwarf planets will be observed by both space- and ground-based telescopes. Such planets will be favorable targets for atmospheric characterization using the transit method because they tend to block a larger relative amount of light crossing the face of a small star. Some scientists have even claimed red dwarf systems as likely candidates for detecting astrobiological signatures of extraterrestrial life-forms (*1*, *15*). Leaving aside this possibility of discovering alien life-forms, measuring the compositional diversity of planets around red dwarf stars—the most common type of star in the Milky Way—is important for piecing together the complex puzzle of small planets' formation and evolution. ■

REFERENCES AND NOTES

1. A. L. Shields *et al.*, *Phys. Rep.* **663**, 1 (2016).
2. R. Luque, E. Pallé, *Science* **377**, 1211 (2022).
3. B. J. Fulton, E. A. Petigura, *Astron. J.* **156**, 264 (2018).
4. J. E. Owen, Y. Wu, *Astrophys. J.* **847**, 29 (2017).
5. J. R. Najita *et al.*, *Astrophys. J.* **766**, 134 (2013).
6. A. S. Rivkin *et al.*, in *Asteroids III* (Univ. Arizona Press, 2002), pp. 235–253.
7. F. Nimmo, R. T. Pappalardo, *J. Geophys. Res. Planets* **121**, 1378 (2016).
8. R. Burn *et al.*, *Astron. Astrophys.* **656**, A72 (2021).
9. S. Sabotta *et al.*, *Astron. Astrophys.* **653**, A114 (2021).
10. M. Schlecker *et al.*, *Astron. Astrophys.* **656**, A71 (2021).
11. C. M. O'D. Alexander, *Phil. Trans. R. Soc. A* **375**, 20150384 (2017).
12. R. Cloutier, K. Menou, *Astron. J.* **159**, 211 (2020).
13. V. Van Eylen *et al.*, *Mon. Not. R. Astron. Soc.* **507**, 2154 (2021).
14. A. R. Neil *et al.*, *Astrophys. J.* **933**, 63 (2022).
15. F. Tian, S. Ida, *Nat. Geosci.* **8**, 177 (2015).

ACKNOWLEDGMENTS

J. Teske thanks X. S. Wang and T. Lichtenberg for helpful discussions that informed this article.

10.1126/science.add7175

POLICY FORUM

HEALTH DISPARITIES

Exclusion cycles: Reinforcing disparities in medicine

Clinical practice, data collection, and medical AI constitute self-reinforcing and interacting cycles of exclusion

By **Ana Bracic[1], Shawneequa L. Callier[2,3], W. Nicholson Price II[4,5]**

Social exclusion of minoritized populations is a pernicious and intractable problem across different domains, from politics to medical practice. In medicine, substantial disparities exist in both experiences and health outcomes for minoritized populations, with origins in systemic racism, implicit bias, historical practice, and social determinants of health (*1*, *2*). We draw on a theory of "exclusion cycles" developed in the context of nonmedical social interactions (*3*) to link known dominant-group and minoritized-group behaviors and demonstrate their self-reinforcing interactions. Interlinked cycles help reveal why exclusion and racial disparities are so intractable in medicine, despite efforts to reduce them on the part of physicians and health systems through strategies focused on individual parts of the cycle such as diverse workforce recruitment or implicit bias training (*1*). This framework highlights particular dangers that may arise through expanding use of big data and artificial intelligence (AI)–based systems in medicine, making bias especially intractable unless tackled directly and early.

In the prototypical exclusion cycle, antiminority culture gives rise to discrimination by some members of the dominant group. Navigating spaces that discrimination has constrained, some members of the minoritized group respond by developing strategies such as withdrawal and self-advocacy. The dominant group misunderstands and may resent many of these strategies, assuming that they are inherent to the minoritized group rather than an outcome of the dominant group's discriminatory behavior, which reinforces antiminority culture and feeds discrimination further.

The exclusion cycle theory, which emphasizes agency on the part of both dominant and minoritized group members, maps onto problematic dynamics in medical practice and research recruitment. We apply our discussion of exclusion cycles to minoritized patients—and we use the term "minoritized" deliberately. Rather than being a minority organically, people are often marginalized by others and consequently minoritized by those with greater social power; individuals may subsequently be treated as minoritized even if successful despite imbalanced power dynamics. Although our principal examples come from the extensive empirical work on the experience of Black patients and research participants in the United States, similar exclusion cycles can be found for many groups, whether minoritized on the basis of race, ethnicity, gender identity, disability, some other marker, or a combination of them. Exclusion cycles transcend specific identities and have the potential to cause pervasive problems throughout medical practice and research.

EXCLUSION CYCLES IN MEDICAL PRACTICE

The clinical encounter

Patient-provider interactions show self-reinforcing exclusion cycles in action (*4*) (see the figure). Medicine has pervasive aspects of antiminority culture. Among the best studied is perception and treatment of pain in Black patients (*5*). Black patients are frequently believed to feel pain less severely, often based on the belief that Black patients are biologically different from white patients (antiminority culture). Accordingly, physicians are more likely to prescribe inadequate doses of pain treatments (discrimination) (*5*). Such poor and discriminatory treatment may lead Black patients to withdraw from treatment relationships or engage in self-advocacy in future encounters with physicians (response strategies) (*4*). And physicians may readily attribute such responses to characteristics of the patients themselves, such as minoritized patients being distrustful or noncompliant (*4*), rather than interpreting them as justifiable responses to discrimination (attribution error). This error leads to a stronger antiminority culture and a refreshed cycle of exclusion. Although each separate stage of the cycle has been well characterized, even sophisticated recent treatments of clinical bias do not link the stages into the self-reinforcing cycle (*6*).

Research participation and recruitment

The dynamics of research participation, including those related to big data and AI, can display their own exclusion cycles. Antiminority culture appears in entrenched perceptions of minoritized patients as "inherently mistrustful" and thus less interested in research participation (*7*). This helps enable discriminatory recruitment practices, such as lessened efforts to recruit minoritized patients directly or reliance on convenient but demographically biased samples (*8*). In one big-data study, a putatively race-neutral convenience sample of perioperative patients was in fact 86.7% white, substantially less diverse than the overall hospital system's 69.8% white patient population (*8*). Minoritized patients, recognizing the discriminatory nature of lessened research engagement, may be justifiably reluctant to participate in research studies, not least because of the demonstrated lack of trustworthiness on the part of institutions (*9*); in the same big-data study, Black patients declined research participation at nearly twice the rate of white patients (*8*). Finally, big-data researchers may erroneously attribute lessened patient participation not as a justifiable reaction to discriminatory practices, but rather as an inherent attribute of minoritized patients—contributing to and strengthening antiminority culture.

THE IMPACT OF BIG DATA AND AI

The use of large-scale datasets and sophisticated machine learning techniques promises positive transformation for the health system, including by improving diagnoses, monitoring patients, and improving quality. Nevertheless, big data and AI show disturbingly similar dynamics of exclusion (see the figure)—or surreptitious and problematic inclusion in undesirable practices (*9*)—but with the added challenges of automation, scale, and the deceptive appearance of objectivity. Indeed, AI exclusion cycles are especially pernicious because of prevalent views that AI systems can decrease bias

[1]Department of Political Science, Michigan State University, East Lansing, MI, USA. [2]Department of Clinical Research and Leadership, George Washington University School of Medicine and Health Sciences, Washington, DC, USA. [3]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [4]University of Michigan Law School, Ann Arbor, MI, USA. [5]Centre for Advanced Studies in Biomedical Innovation Law, University of Copenhagen, Copenhagen, Denmark. Email: bracic@msu.edu, wnp@umich.edu

over time as they learn from patient experience or decrease physician bias by providing a more objective view (*10*). We explain how AI systems might instead entrench bias through exclusion cycles.

First, antiminority culture. AI systems themselves cannot have negative views of minoritized groups. But the humans who write, validate, and deploy AI may be racist or biased, especially given coders' lack of diversity, leading to systems that incorporate antiminority culture (*11*). Even if AI systems are designed by unbiased coders striving for neutrality, those systems derive data from and exist within a medical system that has its own antiminority culture; those views are embedded in the patterns that AI learns. Not only do training data used to develop AI reflect a biased system, but they also typically inadequately represent minoritized patients, based in part on insufficient outreach and the inclusion challenges highlighted above. Perniciously, antiminority culture embedded in data hides beneath a veneer of mechanical objectivity, which may lead to erroneous conclusions that special efforts at inclusion and equity are unwarranted, or that merely increasing dataset size will deal with problems of bias (*10*).

Second, discrimination. AI systems can readily embody bias and discrimination. AI based on biased and incomplete data is unfortunately likely to be less accurate about understudied patients, providing lower-quality recommendations and analyses for those patients (*8*, *12*). For instance, algorithms trained to detect cancerous skin lesions may perform worse on patients with darker skin, because the training datasets principally comprise data from light-skinned patients (*13*). When minoritized patients are included in underlying datasets, AI systems are likely to reflect bias in their recommendations. For instance, an AI system used to recommend follow-up coordination for patients at high risk of medical complications reflected precisely this bias: Because it was trained on underlying medical practice that provided less care to Black patients, the algorithm predicted lower risk for Black patients than similarly situated white patients—and consequently was less likely to recommend interventions (*14*). Notably, these biased recommendations do not stay relegated to computer systems; instead, they feed back into the behavior of physicians, affecting the cycle of exclusion in human-based medical practice. That is, the feedback of the exclusion cycle in data strengthens the exclusion cycle in care.

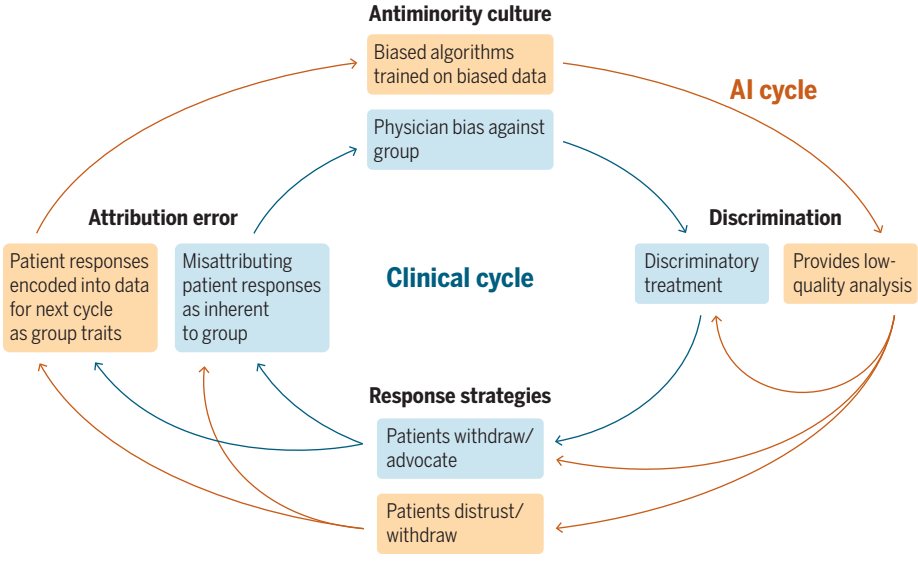
Third, response strategies to negotiate space constrained by discrimination. Members of minoritized communities, facing a demonstrably exclusionary medical system, may be understandably less interested in participating in big-data research (*8*, *9*). Underrepresentation of groups in big datasets has been most thoroughly documented in the context of genomic databases and genome-wide association studies (*12*). But underrepresentation in other big datasets is also pervasive; indeed, a key goal of the US National Institutes of Health "All of Us" initiative is to develop a more representative dataset for precision health (with some notable success). How minoritized patients will respond to the use of AI remains uncertain (*10*); these systems are still being deployed, and it is too early to know whether differences in the quality of recommendations will lead to justifiable differences in trust and adherence. However, as long as discriminatory or poorly performing recommendations exist and are coupled with ensuing withdrawal or resistance strategies, AI's potential to democratize excellence in care is impeded.

Fourth and finally, the attribution error. Minoritized group status is an easy proxy for harder-to-observe attributes. Many outcomes vary substantially by race (*2*). Much of this variance is due to a system with substantial embedded racism and racial bias (*2*), and some of it may be due to justifiable minoritized community response strategies adopted in response to that system. The outcomes that AI "sees" in data will likely vary based on these underlying realities—and be augmented by biases and response strategies in data collection. In a learning AI system, if poor recommendations derived from poor initial data result in justifiably poor adherence to AI recommendations as its own response strategy, future iterations would be even more biased and perform even worse. Putting all this together, it would be surprising if AI did not attribute differences in patient outcomes, and correspondingly in diagnoses or recommendations, to minoritized status (writ broadly). AI systems are thus likely to commit the attribution error themselves, at least implicitly, and use race as a problematic proxy, directly or indirectly—leading to the next bout of antiminority culture embedded within the data and system and strengthening the cycle anew.

AI systems will not always reinforce systemic and data biases. Indeed, AI systems can be used to diagnose bias embedded within apprenticeship-like medical training and other existing medical practices (*15*). Potentially, AI systems using unsupervised learning could even identify and account for underlying disparities, improving equity in care (*10*). However, given the self-reinforcing nature of AI-involved exclusion cycles, a long history of bias in medical care (*1*, *2*), and disparities in data collected, the potential negative impacts demand focus even against the positive possibilities of AI.

Medical practice and AI create overlapping exclusion cycles

Around a generalized structure of the four-step cycle, we depict two interacting exclusion cycles: clinical encounters (blue) and artificial intelligence (AI) products that result from, and influence, data surrounding those clinical encounters (orange). Each cycle self-perpetuates, but the cycles also interact at various points.



GRAPHIC: K. FRANKLIN/*SCIENCE*

POLICY IMPLICATIONS

Although the pieces of exclusion cycles in clinical encounters and AI and big data may

already be known, connecting the pieces explicitly into exclusion cycles has three principal implications: First, exclusion can be self-reinforcing, including in AI; second, exclusion cycles between medical practice and AI can interweave and reinforce one another; and third, effective policy interventions will need to take these self- and other-reinforcing dynamics into account.

At the most basic level, the dynamics of exclusion in medicine are self-reinforcing. Racial bias embedded in medical practice and research data collection self-perpetuates by leading to discrimination, which may lead to strategies such as resistance and withdrawal by minoritized patients, and, consequently, to physicians' erroneous attribution of problems to those patients' inherent characteristics. In AI, this cycle also exists, with the deceptive shield of algorithmic objectivity. Machines cannot be racist, but they can and do participate in self-reinforcing exclusion cycles, based both on the involvement of imperfect humans and training on data that reflect and embody systemic racism. Given this initial involvement, systemic bias will be perpetuated and reinforced through the dynamics of big data and AI themselves. There is no reason to expect improvement over time (*10*), absent active intervention.

Crucially, exclusion cycles in medicine are also other-reinforcing. Even putatively unbiased AI systems learn bias from biased training data and return biased recommendations—but similarly, human physicians, nurses, and other care providers, even if unbiased themselves, will receive biased recommendations and analyses from such AI systems. Disturbingly, we should then expect that the addition of bias from AI systems would itself then be reinforced within care-based exclusion cycles.

Imagine a perfectly unbiased physician in an unbiased care system that imports a highly touted AI system trained in a real-world biased health care system to improve patient pain care. The AI system has learned from biased data to recommend inadequate doses of medication to Black patients and makes those recommendations to the unbiased physician. If clinicians dutifully take the advice from the AI system, prescribing inadequate doses, the system has essentially replicated an antiminority culture, driven discriminatory treatment, and set the stage for understandable patient response strategies and physician attribution errors. After enough cycles of this, the care exclusion cycle can continue even without the AI intervention. Skill fade and automation bias—in which providers rely on automation, including AI, and gradually lose expertise in, e.g., racially equitable provision of care—may augment this process. Though individual-stage dynamics have been noted—biased AI can obviously lead to biased care (*13*)—the ability of entire self-reinforcing cycles to reinforce one another has gone unrecognized.

From a policy intervention standpoint, tackling bias in either system independently is thus woefully insufficient. Bias in either the human care cycle or the AI–data overlay can reintroduce bias at the other level, even positing the absence of bias in human actors themselves. Bias is a systemic infection that cannot be treated in only one place; it must be treated systemically. Accordingly, efforts to decrease bias in medical care at the physician level must be coordinated with (i) efforts to ensure that algorithms trained on existing data do not themselves incorporate biases, and (ii) efforts to guard against the introduction of self-reinforcing biases in algorithms that are deployed into the care process—in contrast to existing proposals that tend to treat problems of care and AI bias largely separately (*6*, *8*, *13*). Research into best practices for coordination and de-biasing interventions could help, as could placing minoritized physicians and data scientists on care and research teams.

Other interventions may best be targeted at different exclusion cycle stages. Interventions focused on algorithms, whether during regulatory review or other governance efforts, are likely best focused on discriminatory outputs—which can be most readily measured—and on attribution errors, especially for explainable AI systems. That is, breaking the exclusion cycle for algorithms is likely easiest by identifying discriminatory outcomes and ensuring that system learning about performance does not problematically incorporate new biases about patient characteristics. Carefully evaluating performance data in the process of retraining could help break the propagation of exclusion cycles—for instance, by distinguishing attribution errors (new learnings about minoritized patients) from decreased adherence to poor (or even accurate) predictions. Antiminority culture is tougher to address, and patient-focused strategies are ethically inappropriate to tackle without first addressing other parts of the cycle. Attempting to promote minoritized group trust in current biased systems is more likely to understandably backfire than to help fix the cycle. Rather, minoritized community strategies to combat systemic constraints will hopefully evolve as the system improves, amid necessary efforts to demonstrate institutional trustworthiness to minoritized communities.

By contrast, interventions for physicians can more readily focus on antiminority culture (through education on systemic biases, including in AI) and discrimination (through monitoring and evaluation of disparate treatment). A rich literature describes such potential interventions (*1*, *2*, *4*).

The interactions that we describe here occur within the context of broader systemic racism in health care (*1, 2, 4, 5*). We focus here on big data and AI, relatively new systemic elements, but patient and provider experiences—and interventions to improve them—are also profoundly influenced by barriers to access, proximity to care, language barriers, the hidden curriculum of medicine and its depictions of minoritized groups, and other factors. Increasing equity must include interventions across the health care system (*1, 2*). Although we cannot address the full scope of systemic racism here, we suspect the exclusion cycle framework will provide a useful tool in analyzing and addressing other aspects of systemic racism. Future empirical research could analyze the details of exclusion cycles in medicine and elsewhere.

The particular interventions we mention here are known. But the combination is important; these issues cannot be addressed seriatim but must be faced in coordination. Though the addition of big data and AI to medicine promises substantial gains, they complicate the picture for reducing bias and require careful efforts to ensure that progress on one front is not rapidly lost on another. ■

REFERENCES AND NOTES

1. M. A. Fair, S. B. Johnson, *Science* **372**, 348 (2021).
2. Institute of Medicine, Unequal Treatment; Confronting Racial and Ethnic Disparities in Healthcare (National Academies Press, 2003).
3. A. Bracic, *Breaking the Exclusion Cycle: How to Promote Cooperation Between Majority and Minority Ethnic Groups* (Oxford Univ. Press, 2020).
4. D. B. Matthew, *Just Medicine: A Cure for Racial Inequality in American Health Care* (NYU Press, 2018).
5. K. M. Hoffman, S. Trawalter, J. R. Axt, M. N. Oliver, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4296 (2016).
6. M. Sun, T. Oliwa, M. E. Peek, E. L. Tung, *Health Aff.* (Millwood) **41**, 203 (2022).
7. L. M. Crawley, *J. Natl. Med. Assoc.* **93** (Suppl), 14S (2001).
8. K. Spector-Bagdady *et al.*, *Health Aff.* (Millwood) **40**, 1892 (2021).
9. S. Callier, S. M. Fullerton, *J. Law Med. Ethics* **48** (S1), 115 (2020).
10. B. Babic, S. Gerke, T. Evgeniou, I. G. Cohen, *Science* **366**, 1202 (2019).
11. R. Benjamin, *Soc. Forces* **98**, 1 (2020).
12. C. N. Rotimi, S. L. Callier, A. R. Bentley, *Science* **371**, 565 (2021).
13. A. S. Adamson, A. Smith, *JAMA Dermatol.* **154**, 1247 (2018).
14. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, *Science* **366**, 447 (2019).
15. E. Pierson, D. M. Cutler, J. Leskovec, S. Mullainathan, Z. Obermeyer, *Nat. Med.* **27**, 136 (2021).

ACKNOWLEDGMENTS

W.N.P. is supported by Novo Nordisk Foundation grant NNF17SA0027784. We thank S. Bender and P. Bugga for research assistance and E. Trapecar for design assistance.

10.1126/science.abo2788

Mentees can benefit from mentors who share the same race or gender.

PHOTO: JOHNNYGREIG/GETTY IMAGES

PSYCHOLOGY

Connecting for success

Interventions that increase our sense of belonging can help us confront new challenges

By **Matthew Lieberman**

What if there was a pill that could significantly improve your child's grade point average (GPA) and increase their long-term success without negative side effects? What would you pay for it, and how might it achieve this seemingly magical outcome—by increasing IQ, by enhancing willpower? What if I told you there was a way to accomplish the same ends without affecting either of these variables and without the help of a fictional pharmaceutical intervention? As unlikely as it sounds, the key to raising students' GPA often starts with raising their sense of belonging at school.

All mammals are born immature and incapable of caring for themselves, so they have evolved a need for social connection and some capacity for empathy. From an evolutionary perspective, the social pain infants feel when separated from their caregiver and the distress that this causes in caregivers keeps the two groups bonded long enough for the infant to mature and reach self-sufficiency. Humans, unlike many other species, however, have a generalized need to feel like we belong in groups, a need that flies in the face of the American myth of rugged individualism, making it all the more important to come to terms with.

Belonging, by social psychologist Geoffrey Cohen, focuses on the current "crisis of belonging" preventing many teens and young adults from reaching their full potential. In the book, Cohen alternates between telling the deep history of successful social psychological interventions and focusing on modern interventions that are being used to astounding effect. It is in the description of these more recent interventions that the book shines.

Belonging
Geoffrey L. Cohen
Norton, 2022.
448 pp.

The main tool under consideration is what Cohen and his colleagues somewhat dubiously call "wise interventions." Wise interventions are best understood in contrast to the better-known "nudges." Nudges depend on people's mental laziness. For example, health care choice forms with precompleted options count on people sticking with the default selections. In contrast, wise interventions focus on "subjective construals"—our idiosyncratic understandings of our situations and ourselves (*1*).

Sometimes a modest change to how we understand a situation can have dramatic cascading effects over time. Believing that you are up to a challenge can alter how you respond to the next roadblock you face, which in turn affects your ability to overcome the one after that. What makes wise interventions more remarkable is that they are typically inexpensive, simple, and not zero-sum. Helping members of one group feel that they belong does not deprive other groups of any resources or opportunities.

The wise interventions discussed in the book take a variety of forms, from small gestures of welcome, to behaviors that affirm a person's values, to opportunities that allow an individual to reinterpret evidence. They can even take the form of working with a mentor who is of the same race or gender, which can help a mentee to more easily imagine surmounting comparable challenges.

Consider Cohen's 2011 *Science* paper, written with Greg Walton, which showed that a brief wise intervention with first-year Black college students significantly closed a race-based GPA gap 3 years later (*2*). The intervention? Allowing Black students to see that white students experienced uncertainty about whether they belonged at their respective institutions early on in their college careers—an act that helped the students realize that others felt this way regardless of their race. This hour-long intervention at the start of college also produced demonstrable effects on participants' work success and general well-being a decade later.

No book is perfect. Some of the time that Cohen devotes to the well-trod history of social psychology might have been better spent on why we have a need for belonging in the first place, and some of the chapters seem tangential to the book's thesis. And although Cohen rightfully promotes the importance of situational factors in academic performance, he unnecessarily denigrates the reality of individual differences in ability. We should nurture both belonging and ability, which is harder when school districts increasingly deny the latter to support the former.

But ultimately, these issues are quibbles. If we want more people to have the chance to maximize their abilities and their contributions to society, reading *Belonging* is the right place to start. ■

The reviewer is at the Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095, USA, and is the founder of Resonance Inc., Los Angeles, CA 94134, USA. Email: lieber@ucla.edu

REFERENCES AND NOTES

1. M. D. Lieberman, *Psychol. Rev.* **129**, 830 (2022).
2. G. M. Walton, G. L. Cohen, *Science* **331**, 1447 (2011).

10.1126/science.add2900

MARINE ECOLOGY

A tale of troubled waters

Vietnamese refugees faced misdirected rage for a marine ecosystem ravaged by industry, argues a journalist

By **Mary Ellen Hannibal**

Pity the Gulf of Mexico. This geological wonder hosts fantastically deep reserves of oil, to which we have responded by piercing its shoreline with thousands of wells. The landscape is also dotted with industrial operations that transform its oil into plastic bags and toothbrushes, among other items of our everyday comfort. Not only does the region's biodiversity suffer from regular oil spills (the most frequent in the world) and the release of noxious chemicals into air, water, and earth, people suffer as well. The intersection of politics, history, industrial pollution, and ugly human behavior in these waters is the focus of Kirk Wallace Johnson's new book, *The Fishermen and the Dragon*.

Johnson begins his narrative in the 1970s, introducing readers to hotheaded fisherman Billy Joe Aplin, who struggles to make a living in the waters near Seadrift, Texas, a tiny community between Galveston and Corpus Christi. Pulling deformed crabs from inadequate catch, Aplin reasons that both must result from toxic waste in the water courtesy of a massive aluminum smelting plant operated by Aluminum Company of America (Alcoa) nearby. He tries to organize his fellow fishermen to challenge the industrial giant, but despite their own poor harvests, they greet his entreaties with passivity and indifference. They do rouse to ire and action, however, when Aplin shifts his angry focus to South Vietnamese refugees new to the area.

US officials had resettled many of the displaced South Vietnamese in Louisiana, where they found work at a crab-picking plant. (Local inmates given the chance to put in the "feverish hours" required of this work opted to sit behind bars instead.) Desperate for workers, the owner of a crab-picking plant in Seadrift exhorted the Vietnamese to move to Texas, where, despite high rates of poverty and unemployment, locals would not do the work. The Vietnamese would, and they did.

Drawn to life on the water, which many had left behind in Vietnam, the refugees gravitated to fishing themselves. Their record catches drew the ire of Americans on the water, who worked fewer hours with far less efficiency. The local white community soon called in the Ku Klux Klan (KKK) to deliver fresh horror to the traumatized Vietnamese. Most of Johnson's book is taken up with the gruesome consequences of this white supremacy, which created havoc and destroyed lives and livelihoods, but a fascinating back story—the saga of the Southern Poverty Law Center (SPLC)—emerges as the Klan's antics attract media attention and then outrage.

Founded in 1971 by Morris Dees, the SPLC charged into the Gulf Coast melee to defend the Vietnamese against white supremacy. Johnson dramatizes Dees's evolution from "backcountry hayseed" to sophisticated advocate. His litigation on behalf of the Vietnamese is exalted, inspiring, and successful.

And yet, by the end of the book, the SPLC is beset by controversy. Dees is cited for sexual harassment and fired from the organization he founded. Equally troubling is the charge that the big bucks his highly successful marketing brought in were not always directed to help redress the United States' wrongs.

The second major thread of Johnson's book focuses on Diane Wilson, a Seadrift local with many more generations behind her than Aplin. The malfeasance in this section of the book issues not from vindictive white people asserting racial purity but from sophisticated lawyers and business people asserting profit above all peril.

A mother of four, Wilson is poor and married to a traumatized Vietnam vet. She scrapes together something close to a living by working on the water, and, like Aplin, she begins to associate anomalies in the region's sea life with nearby industry. A true citizen scientist, over the years, Wilson helps collect more than 30 million "nurdles"—small pellets from which most plastic products are created, which eventually convey industrial chemicals, pesticides, toxins, and dangerous bacteria into the food chain—from the shoreline. Wilson undergoes several hunger strikes to force the corporate hand, and in 2019 she succeeds in winning a $50 million settlement from Formosa Plastics.

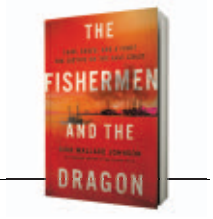
Johnson's twin stories come together tangentially. They both involve cruelty and destruction in the Gulf of Mexico, but the KKK predates the capitalist forces that have exploited and polluted the region. Both the Klan and industry have wrought terrible destruction on people and ecosystems, but any direct connection between these evils is not in evidence here.

This book does demonstrate, however, that humanity's propensity for immorality, greed, and laziness can come from very different directions and in very different forms. Thankfully, our more admirable qualities often manifest in myriad ways and from unexpected sources as well. ■

10.1126/science.ade0604

The reviewer is the author of *Citizen Scientist: Searching for Heroes and Hope in an Age of Extinction* (The Experiment, 2016). Email: maryellenhannibal@gmail.com

The Fishermen and the Dragon: Fear, Greed, and a Fight for Justice on the Gulf Coast
Kirk Wallace Johnson
Viking, 2022. 384 pp.



Members of the Ku Klux Klan set fire to the "USS Viet Cong" in Santa Fe, Texas, on 14 February 1981.

PHOTO: ED KOLENOVSKY/AP PHOTO

China has taken steps to protect the South China Sea's coral reef species, such as the giant clam.

PHOTO: YOUFANG SUN

Edited by **Jennifer Sills**

Joint protection of a crucial reef ecosystem

The South China Sea, situated just north of the Coral Triangle, contains abundant reef bioresources and provides a range of important ecological services (*1*). However, South China Sea reefs have suffered serious degradation due to local human activities, ocean warming, and outbreaks of coral-eating crown-of-thorns starfish (*2*, *3*). Furthermore, overfishing in the South China Sea has decimated its bioresources (*3*). Such degradation impairs ecosystem function and stability and, as a result, ecological goods and services (*4*). Countries adjacent to the South China Sea can act together to mitigate reef ecosystem damage.

Countries in the South China Sea region have been independently expanding their reef protection efforts. In 2021, China's List of State-Protected Wildlife under State Priority Conservation prohibited exploitation of stony corals, giant clams, the giant triton snail (*Charonia tritonis*), and other reef animals such as sea turtles and the humphead wrasse (*Cheilinus undulatus*) (*5*). In addition, China has implemented seasonal fishery closures, designated marine protected areas (*6*), and carried out the removal of crown-of-thorns starfish (*2*). Other countries have also taken steps to conserve coral reefs by establishing marine protected areas, especially Malaysia (*7*) and Thailand (*8*). Fishery closures have been established in the Philippines and Vietnam (*9*). However, the jurisdictional and policy complexities of the South China Sea have made effective synchronous implementation of these protections extremely difficult, which may contribute to illegal, unreported, and unregulated (over-) fishing (*10*), fishing disputes and conflicts (*10*) and illegal trade of protected wildlife (*11*).

To ensure that efforts toward sustainable reef protection are successful, countries bordering the South China Sea should prioritize collaborative, multilateral conservation. In addition to strictly implementing national protection policies, countries should strengthen regional communication and exchanges to promote the consistency and effectiveness of policies. For example, seasonal fishery closures could be enforced collaboratively. Countries could join together to fight against illegal trade of protected reef animals. International cooperation, spearheaded by local countries, could establish early warning and information-sharing systems to control crown-of-thorns starfish. Robust collaboration and consistent policy enforcement among countries adjacent to the South China Sea could make effective coral reef conservation possible.

Youfang Sun[1], Lintao Huang[1], Laurence J. McCook[3,4], Hui Huang[1,2]*

[1]Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China. [2]Sanya National Marine Ecosystem Research Station, Tropical Marine Biological Research Station in Hainan, Chinese Academy of Sciences, Sanya, China. [3]World Wide Fund for Nature Hong Kong, Hong Kong, China. [4]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD, Australia.
*Corresponding author.
Email: huanghui@scsio.ac.cn

REFERENCES AND NOTES

1. M. D. Spalding, C. Ravilious, E. P. Green, *World Atlas of Coral Reefs* (University of California Press, Oakland, 2001).
2. Y. Li *et al.*, *Chin. Sci. Bull.* **64**, 3478 (2019) [in Chinese].
3. T. P. Hughes *et al.*, *Conserv. Biol.* **27**, 261 (2013).
4. F. Moberg *et al.*, *Ecol. Econ.* **29**, 215 (1999).
5. Government of China, "The list of state-protected wildlife under state priority conservation," (2021); www.gov.cn/xinwen/2021-02/09/5586227/files/e007df5cdb364bcdbcb89d169047d6c5.pdf [in Chinese].
6. L. J. McCook *et al.*, *Aquat. Conserv. Mar. Freshw. Ecosyst.* **29**, 33 (2019).
7. Z. bin Talib *et al.*, in *Proceedings of the 1st Regional Workshop on Enhancing Coastal Resources* (2004), pp. 129–136.
8. N. J. Bennett *et al.*, *Mar. Pol.* **44**, 107 (2014).
9. C. D. Elvidge *et al.*, *Front. Mar. Sci.* **5**, (2018).
10. H. Zhang, *J. Boundary Ocean Stud.* **3**, 36 (2018) [in Chinese].
11. V. Nijman, *Biodivers. Conserv.* **19**, 1101 (2010).

10.1126/science.abo0166

Adaptive wildfire mitigation approaches

Wildfires have been increasing in frequency, severity, and intensity in many

landscapes worldwide. The trends are driven by climate change (*1*), changes in land management (*2*) and land cover (*3*), logging (*4*), and ineffective fire mitigation actions (*5*). As a result, extensive parts of some landscapes are now extremely flammable and will remain so for decades to come (*6*). Improving fire management, especially early detection, will require new technologies and strategies (*7*, *8*).

Earlier fire detection, which allows fire crews to arrive sooner and work to contain and suppress a fire at an early stage (*9*), is a promising strategy. Mathematical modeling of lightning strikes can help predict where lightning is most likely to occur and lead to ignitions (*10*). Drone fleets equipped with infrared sensors can confirm whether those ignitions have occurred (*8*). Ground-based sensor networks can be deployed to detect ignitions in areas of high ecological value (*11*). Cameras on fire towers can replace or supplement fire observers, providing round-the-clock fire detection capability, and satellites can inspect vast areas (*8*). The relative flammability of ecosystems can be tracked through a combination of technologies that measure the moisture content of vegetation and new-generation fire models that account for fuel moisture along with vegetation structure and composition (*12*).

Immediate identification of fires increases containment options. In some cases, GPS-guided, low-cost, uncrewed autonomous vehicles can rapidly access ignition points and extinguish fires within minutes (*8*). Aerially dispensed water gliders—one example of such autonomous vehicles—can be deployed from high altitudes in poor weather conditions, providing an alternative when it is too dangerous to deploy aircraft and helicopters (*8*).

Rapid suppression technologies are most effective when fire behavior is mild. Therefore, reducing flammability can facilitate effective rapid suppression. In many ecosystems, decreasing flammability requires either (i) massively increasing deliberate human disturbances such as prescribed burning or (ii) enabling natural processes such as self-thinning and vegetation maturation to facilitate a transition to a less-flammable state (*5*). For forests recovering from logging, fire, or past clearing, strategies to reduce flammability should depend on local conditions, such as proximity to dwellings. Together, minimizing flammability and implementing multiple early detection and containment technologies can contribute to effective firefighting in increasingly fire-prone landscapes.

David Lindenmayer[1]*, Philip Zylstra[2,3], Marta Yebra[1,4,5]
[1]Fenner School of Environment & Society, The Australian National University, Canberra, ACT 2601, Australia. [2]School of Molecular and Life Sciences, Curtin University, Perth, WA, Australia. [3]Evolution and Ecology Research Centre, University of New South Wales, Sydney, NSW, Australia. [4]School of Engineering, The Australian National University, Canberra, ACT, Australia. [5]ANU-Optus Bushfire Research Centre of Excellence, The Australian National University, Canberra, ACT 2600, Australia.
*Corresponding author.
Email: david.lindenmayer@anu.edu.au

REFERENCES AND NOTES

1. M. W. Jones *et al.*, "Climate change increases the risk of wildfires," *ScienceBrief Review* (2020); https://sciencebrief.org/briefs/wildfires.
2. G. M. Jones *et al.*, *Front. Ecol. Environ.* **20**, 210 (2022).
3. P. J. Zylstra, *Austral Ecol.* **43**, 578 (2018).
4. N. Wilson, R. Bradstock, M. Bedward, *Agricult. For. Entomol.* **323**, 109077 (2022).
5. P. Zylstra, D. Bradshaw, D. B. Lindenmayer, *Environ. Res. Lett.* **17**, 044022 (2022).
6. J. M., Furlaud *et al.*, *For. Ecol. Manag.* **502**, 119724 (2021).
7. J. J. Roldán-Gómez, E. González-Gironda, A. Barrientos, *Appl. Sci.* **11**, 363 (2021).
8. M. Yebra *et al.*, *Austral. J. Emerg. Manag.* **76**, 4 (2021).
9. M. Plucinski, *For. Sci.* **58**, 390 (2012).
10. R. Coughlan *et al.*, *Meteorol. Appl.* **28**, e1973 (2021).
11. U. Dampage *et al.*, *Sci. Rep.* **12**, 46 (2022).
12. K. Rao, A. P. Williams, A. G. Konings, "Remotely sensed live fuel moisture content across the western USA and its relationship with climate aridity," American Geophysical Union (2019); https://ui.adsabs.harvard.edu/abs/2019AGUFMNH52A..03R/abstract.

10.1126/science.ade4721

Restore China's coastline from the ground up

In China, decades of industrialization and growing cities have replaced coastal mudflats, sandy beaches, mangroves, and rocky reefs with engineered structures such as seawalls, break walls, and dikes (*1*–*3*). The loss of coastal wetlands exposes coastal cities and communities to substantial damages from storms (*4*) and extreme events, which are increasing in frequency, intensity, and scale (*5*). The degradation of coastal ecosystems also threatens migratory birds, which depend on the wetland habitats along their flyway (*6*). To conserve and restore natural coastlines, policies must prioritize quantitative local assessments of successes and failures and use that information to guide national and regional priorities.

National policies issued since 2008 have had some success in stabilizing the retention rate of China's natural coastline. By the end of 2018, China had invested about US$1.5 billion, restored 156.6 km of coastline, and recovered 3549 ha of coastal wetlands (*7*). However, the lack of effective and timely assessments of local ecological environments also resulted in failures: The non-native species that were introduced to slow erosion, such as common cordgrass (*Spartina anglica)* and smooth cordgrass (*Spartina alterniflora*), led to damaging biological invasions (*8*). Attempts to implement national goals without considering the effects on local species undermined restoration goals.

Quantifying coastal losses at the local level will clarify how national and regional strategies can remain sensitive to local conditions. Instead of providing one-size-fits-all goals that all local jurisdictions try to achieve, national and regional governments should craft plans that incorporate consultation with local experts and adaptation to each location's needs. Flexibility to respond to local pressures will be particularly important in the face of mounting challenges such as climate change.

Local monitoring and feedback linked to management action, which previous policies have lacked, are also vital to guiding policies toward desired impacts. Monitoring of China's coastline changes, as well as environmental and human drivers such as sea-level rise, erosion-accretion, and human population pressures, should be conducted in every coastal region. The data should be collected and analyzed regularly, and corrections to the plan should be made quickly using big data and decision-making tools. With national and regional priorities as a framework, restoration should be based on each local ecosystem's structure, functions, composition, and services.

Xuege Wang[1,2,3], Fengqin Yan[2,4], Vincent Lyne[5], Fenzhen Su[1,2,3,4]*
[1]School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China. [2]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. [3]Collaborative Innovation Center of South China Sea Studies, Nanjing University, Nanjing 210093, China. [4]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. [5]Institute for Marine and Antarctic Studies–Hobart, University of Tasmania, Hobart, TAS 7004, Australia.
*Corresponding author. Email: sufz@lreis.ac.cn

REFERENCES AND NOTES

1. H. K. Lotze *et al.*, *Science* **312**, 1806 (2006).
2. X. Wang *et al.*, *Ocean Coast. Manag.* **204**, 105542 (2021).
3. F. Yan *et al.*, *Ecol. Indic.* **116**, 106337 (2020).
4. F. Sun *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5719 (2020).
5. "Extreme weather hits China with massive floods and scorching heat," *The New York Times* (2022).
6. C. Studds *et al.*, *Nat. Commun.* **8**, 14895 (2017).
7. Ministry of Natural Resources of the People's Public of China, "Blue Bay Restoration Report" (2019) [in Chinese].
8. J. Ren *et al.*, *Sci. Adv.* **7**, eabi8943 (2021).

10.1126/science.ade3911

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

EVOLUTIONARY BIOLOGY

Fishy business

Three-spined sticklebacks are nominally marine species, but after the Pleistocene glacial retreat, they colonized freshwater lakes widely. Weber *et al.* have capitalized on the divergent populations that arose to test ideas about the costs of immune responses in animals. Using a combination of wild and laboratory-created hybrids, the authors found that lakes contained fish with differences in immune responsiveness and different susceptibilities to a tapeworm parasite. Fish that mounted an extreme immune response, as measured by fibrosis, showed a marked reduction in fecundity. Immune response–related loci show evidence of opposing selection among resistant and susceptible populations. The ecological conditions that have caused these divergences remain unidentified, but knowledge of the genes involved does indicate a way to investigate immune costs in other species. —CA *Science*, abo3411, this issue, p. 1206

Three-spined sticklebacks (right) can either resist or tolerate tapeworm parasites (left), with consequences for the fish's fecundity.

PHOTO: NATALIE STEINEL

CORRELATED MATERIALS

Understanding the cuprate parent state

Calculating the properties of materials with strong electron correlations, such as cuprate superconductors, from first principles is extremely challenging. Cui *et al*. developed a strategy that combines quantum embedding, ab initio quantum solvers, and periodic quantum chemistry. They used this strategy to establish causal relationships between the material composition of parent compounds of mercury-based cuprate superconductors and their magnetic properties. This approach may also prove fruitful in the more challenging case when the parent compounds are doped to induce superconductivity. —JS

Science, abm2295, this issue p. 1192

SOLID-STATE PHYSICS

Nanowire tip probes magnetism

The material samarium hexaboride (SmB_6) has been predicted to be a topological insulator that also features strong electronic correlations. Topological insulators exhibit so-called spin-momentum locking, which has been put to use in spintronics applications. Aishwarya *et al*. exploited these properties to achieve spin-polarized tunneling through a scanning tunneling microscope (STM) tip. The researchers mounted an SmB_6 nanowire onto a trimmed tungsten tip. This combined tip was then used to image the magnetic state in the material $Fe_{1+x}Te$. Unlike conventional spin-polarized STM, the authors' setup does not require the application of an external magnetic field, which makes for a less-invasive probe. —JS

Science, abj8765, this issue, p. 1218

EL NIÑO

A faunal tipping point

El Niño events occur in the eastern Pacific ocean when warm surface waters prevent normal upwelling of nutrient-rich, cooler water. These events are unpredictable but are increasing in frequency and intensity, likely due to climate change, and have impacts on marine and terrestrial environments, ocean currents, and climate. Broughton *et al*. looked across a 12,000-year record from a coastal site in Baja, California, at the impacts of El Niño events as measured by changes in the fossil record (see the Perspective by Sandweiss and Maasch). They found that in centuries with five or more events, marine fauna decreased, whereas terrestrial diversity increased, indicating a tipping point for these communities. —SNV

Science, abm1033, this issue p. 1202; see also add8890, p. 1153

HYDROFORMYLATION

Taking pressure off cobalt

Hydroformylation is a high-volume chemical process that manufactures aldehydes from olefins, hydrogen, and carbon monoxide. Cobalt carbonyl complexes served as the original catalysts, and although they have been superseded in many cases by rhodium, they are still used with long-chain olefins. However, high carbon monoxide pressure is generally applied on the assumption that the cobalt will otherwise undergo deactivation by clustering. After performing both kinetic and spectroscopic studies, Zhang *et al*. report that cobalt carbonyl

catalysts are stable at lower carbon monoxide pressure than previously thought. Moreover, recently reported phosphine-stabilized cobalt catalysis could not be reproduced. —JSY

Science, abm4465, this issue p. 1223

SOLAR CELLS

Avoiding lithium to fight iodide

Although lithium-doped organic hole transport layers (HTLs) enable efficient charge extraction in perovskite solar cells, they also promote degradation, because lithium ions can adsorb water and the positive radicals that form promote migration of iodide anions from the perovskite layer. Wang *et al*. developed an organic HTL that couples positive polymer radicals and molecular anions through an ion-exchange process. The resulting highly conductive HTL has improved energy-level alignment with the perovskite compared with commonly used lithium-doped HTLs. High thermal and chemical stability with respect to iodide migration resulted in stable perovskite solar cells that maintained 92% power conversion efficiency of 23.9% for 1000 hours at 85°C. —PDS

Science, abq6235, this issue p. 1227

BRAIN STRUCTURE

Uncharted territory in the brain

The ridges and grooves ("sulci") of the human brain have been studied for centuries. However, Willbrand *et al.* report that a small groove, the inframarginal sulcus (*ifrms*), is a previously uncharacterized brain feature. The authors first manually identified *ifrms* using high-resolution neuroimaging in vivo and subsequently in ex vivo brains. The *ifrms* was present in all human samples across a range of ages, but only in about half of chimpanzee brains. A deep-learning approach could automatically locate *ifrms* in individual human brains. Functionally, the *ifrms* may be a landmark for the cognitive control network in humans. —RW

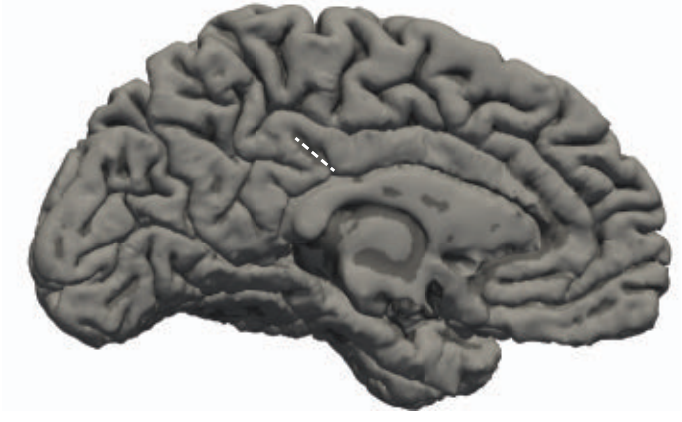
Sci. Adv. 10.1126/sciadv.abn9516 (2022).

HIV

Analyzing antibodies

A challenge for HIV-1 vaccine development is the need to induce a polyclonal neutralizing antibody (nAb) response in vaccine recipients. Saunders *et al.* vaccinated rhesus macaques with a stabilized HIV-1 envelope trimer plus a Toll-like receptor 7/8 agonist with the goal of eliciting such nAbs. The authors observed that vaccinated macaques developed potent nAbs that targeted multiple sites on the envelope, including the CD4-binding site. This treatment translated into protection against mucosal simian-HIV challenge in the macaques, so these results support the further development of this vaccine for humans. —CSM

Sci. Transl. Med. **14**, eabo5598 (2022).



High-resolution brain imaging has revealed a previously unrecognized groove (white dashed line) in the human brain that is not uniformly present in other hominoid species.

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

CANCER IMMUNOLOGY

Melanoma fishing for a complement?

The complement cascade is an arm of the innate immune system that recruits immune cells and forms membrane attack complexes (MACs) to kill target cells. This cascade plays a poorly understood role in immune responses to melanoma. In mice and humans, Liu *et al.* found that complement pathways are activated along blood vessels supplying melanomas. MACs within melanoma tumors were associated with activated neutrophils generating reactive oxygen species and neutrophil extracellular traps. Pharmacological inhibition of complement hampered neutrophil activation and the concomitant vascular permeability of the tumors. Both neutrophil-depleted animals and mice lacking complement factor C6, a critical element in MAC formation, showed reduced tumor dissemination into the lungs. —STS

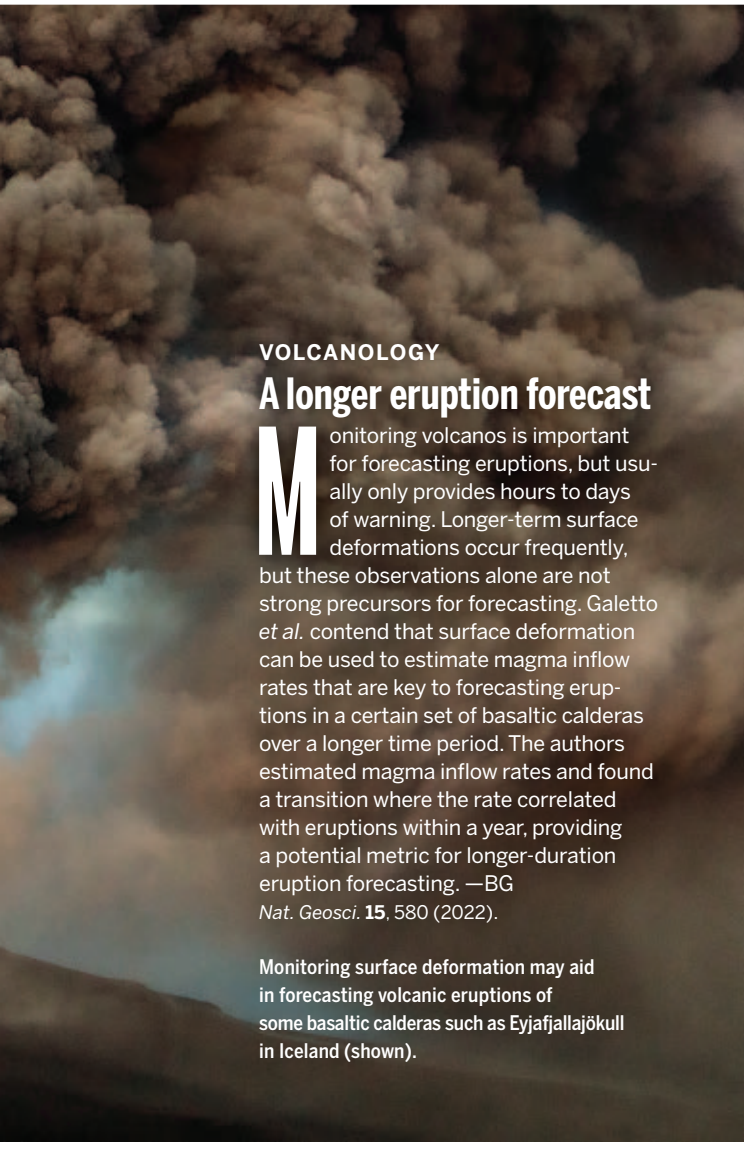
Proc. Natl. Acad. Sci. U.S.A. **119**, e2122716119 (2022).

MICROBIOME

Phage to soothe inflammation

The gut microbiota can exacerbate intestinal inflammation

CREDITS: (LEFT TO RIGHT) WILLBRAND *ET AL.*; ETIENNE DE MALGLAIVE/GETTY IMAGES



VOLCANOLOGY

A longer eruption forecast

Monitoring volcanos is important for forecasting eruptions, but usually only provides hours to days of warning. Longer-term surface deformations occur frequently, but these observations alone are not strong precursors for forecasting. Galetto *et al.* contend that surface deformation can be used to estimate magma inflow rates that are key to forecasting eruptions in a certain set of basaltic calderas over a longer time period. The authors estimated magma inflow rates and found a transition where the rate correlated with eruptions within a year, providing a potential metric for longer-duration eruption forecasting. —BG

Nat. Geosci. **15**, 580 (2022).

Monitoring surface deformation may aid in forecasting volcanic eruptions of some basaltic calderas such as Eyjafjallajökull in Iceland (shown).

in inflammatory bowel disease (IBD), but broad-spectrum antibiotics are not effective at treating it. Federici *et al*. investigated whether bacteriophages, which infect and kill specific bacteria, could be used to treat IBD. In a cohort of 537 patients, the authors found *Klebsiella pneumoniae* strains associated with IBD flare-ups. A consortium of clinical strains of the bacterium was used to inoculate germ-free mice to induce intestinal inflammation. From these, two phages were selected for oral administration into 18 healthy individuals in a phase 1 trial. The treatment was well tolerated and reduced the abundance of *K. pneumoniae* in fecal samples of the treated cohort, showing promise for treatment. —GKA

Cell **185**, 2879 (2022).

GENETICS

Invasion of the killer rabbits

Rabbits introduced into Australia have wreaked vast environmental and economic damage and brought about the extirpation of endemic species. What is it about rabbits that has allowed them to establish a firm foothold in an environment alien to them? In a genetic analysis, Alves *et al.* discovered that by 1870, domesticated rabbits were found around most settlements. Mitochondrial genome sequencing showed that it took the introduction of just a few wild females to allow the species to spread explosively across the continent. Admixture analysis revealed that genetic refreshment of the inbred domestic rabbits by wild-type genes endowed an invasive capacity amplified by subsequent spread across Tasmania and New Zealand. As for the inoculum, Thomas Austin requested rabbits from his family home in Somerset, England, for hunting at Geelong, Victoria. From 24 sent in 1862, Austin had bagged over 20,000 rabbits by 1865. —CA

Proc. Natl. Acad. Sci. U.S.A. **119**, e2122734119 (2022).

METABOLISM

Why cancer cells do fermentation

The Warburg effect, aerobic glycolysis, plays an enigmatic role in cancer. Is it really an effect or is it a cause? Many have puzzled over why cancer cells use aerobic glycolysis, fermenting glucose to lactate, which is relatively inefficient, even when oxygen is present. In human cancer cells in culture, Wang *et al.* found that lactate excretion did not correlate with cell proliferation but happened when the capacity of the electron shuttles produced in glycolysis to move electrons to the mitochondria to regenerate NAD^+ was exceeded. Enhancing shuttle performance by expressing component enzymes decreased lactate production but did not affect cell proliferation. Thus, fermentation does not seem to drive cell proliferation, nor is it the preferred mode of metabolism in cancer cells. —LBR

Mol. Cell. 10.1016/j.molcel.2022.07.007 (2022).

STEM EDUCATION

Thriving through transitions

Community colleges are a critical component of undergraduate STEM education in the United States, with 44% of STEM degree holders attending a community college at some point. Despite this, the experiences of transfer students remain relatively understudied. Gray *et al.* used Schlossberg's model for analyzing human adaptation to transition to understand how STEM community college transfer students navigated and adapted to a new 4-year university. Analyses illumined two broad pathways that students tended to diverge into: thriving, characterized by support and a sense of belonging, and surviving, characterized by stress and barriers. These results provide a framework for understanding the myriad of factors influencing the transfer process as the STEM community works toward more thriving and less surviving among transfer students. —MMc

CBE Life Sci. Ed. 10.1187/cbe.21-09-0261 (2022).

OPTOMECHANICS

Stabilizing optomechanical resonators

The confinement of light to high-quality optical cavities and resonators provides a route to enhanced sensing capabilities and the ability to make precision measurements of targets trapped within the cavity. However, as the light intensities increase and cavity dimensions shrink in size, the light can cause photothermal instabilities of the cavity, adversely affecting the overall performance of the system. Qin *et al.* demonstrate that placing an intracavity window material within the cavity can reduce or even cancel the photothermal instabilities. Insertion of the window material, the photothermal properties of which can be selectable or tunable for different setups, alters the effective photothermal response of the combined system. Mitigating for and removing the photothermal instabilities in such a general and passive way should be useful for a broad range of optical- and photonic-based technologies. —ISO

Optica **9**, 924 (2022).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

MEDICINE

Comparing lingering symptoms

The emergence of Long Covid has increased interest in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), a condition that often follows an infection and has many similarities to Long Covid. In a Perspective, Marshall-Gradisnik and Eaton-Finch discuss what we might learn from the extensive research on ME/CFS. The authors highlight hypotheses that have been proposed to explain the heterogeneous symptoms associated with ME/CFS, which may also apply to Long Covid. They also discuss what is needed to take research forward: clear diagnostic criteria and biomarkers, animal models, and comprehensive clinical trials. Applying the lessons learned from ME/CFS could improve Long Covid research and benefit patients with either condition. —GKA

Science, abo1261, this issue p. 1150

LAND USE CHANGE

Forest loss for food

Agricultural expansion is recognized as a major driver of forest loss in the tropics. However, accurate data on the links between agriculture and tropical deforestation are lacking. Pendrill *et al.* synthesized existing research and datasets to quantify the extent to which tropical deforestation from 2011 to 2015 was associated with agriculture. The authors estimated that at least 90% of deforested land occurred in landscapes where agriculture drove forest loss, but only about half was converted into productive agricultural land. Data availability and trends vary across regions, suggesting complex links between agriculture and forest loss. —BEL

Science, abm9267, this issue p. 1168

MOLECULAR BIOLOGY

A passage through nucleosomes

In the eukaryotic cell nucleus, genomic DNA is wrapped around histones to form nucleosomes, the basic units of the beads-on-a-string structure of chromatin. Although the nucleosomes are physical obstacles for the DNA-transcribing RNA polymerase II, it somehow passes through them while maintaining their structure. Ehara *et al*. obtained structural snapshots of the polymerase passaging through a nucleosome using cryo–electron microscopy. They found that the polymerase forms a huge elongation complex with multiple transcription elongation factors, which mediates the disassembly of the downstream nucleosome and its subsequent reassembly behind the polymerase with the aid of the histone chaperone FACT. —DJ

Science, abp9466, this issue p. 1169

NEURODEVELOPMENT

Neanderthal brain development

Neanderthal brains were similar in size to those of modern humans but differed in shape. What we cannot tell from fossils is how Neanderthal brains might have differed in function or organization of brain layers such as the neocortex. Pinson *et al*. have now analyzed the effect of a single amino acid change in the transketolase-like 1 (TKTL1) protein on the production of basal radial glia, the workhorses that generate much of the neocortex (see the Perspective by Malgrange and Nguyen). Modern humans differ from apes and Neanderthals by this single amino acid change. When placed in organoids or overexpressed in nonhuman brains, the human variant of TKTL1 drove more generation of neuroprogenitors than did the archaic variant. The authors suggest that the modern human has more neocortex to work with than the Neanderthal did. —PJH

Science, abl6422, this issue p. 1170; see also ade4388, p. 1155

CLIMATE CHANGE

Getting tipsy

Climate tipping points are conditions beyond which changes in a part of the climate system become self-perpetuating. These changes may lead to abrupt, irreversible, and dangerous impacts with serious implications for humanity. Armstrong McKay *et al*. present an updated assessment of the most important climate tipping elements and their potential tipping points, including their temperature thresholds, time scales, and impacts. Their analysis indicates that even global warming of 1°C, a threshold that we already have passed, puts us at risk by triggering some tipping points. This finding provides a compelling reason to limit additional warming as much as possible. —HJS

Science, abn7950, this issue p. 1171

DOMESTICATION

Donkeys' African origins

Donkeys have been important to humans for thousands of years, being the primary source of work and transport for many cultures. Unlike horses, little was known about the origin and domestication of donkeys. Todd *et al*. sequenced the genomes of modern and ancient donkeys and found evidence of an eastern African origin over 7000 years ago, with subsequent isolation and divergence of lineages in Africa and Eurasia. The authors also reveal the imprint of desertification on divergence among groups and specifics about donkey breeding and husbandry, including selection for large size and the practice of inbreeding. —SNV

Science, abo3503, this issue p. 1172

MESOSCOPIC PHYSICS

Probing an exotic state

The nature of some fractional quantum Hall (FQH) states can be determined by measuring the thermal conductance of the state's edge. Such measurements have suggested that the well-known FQH state at 5/2 filling, which is thought to harbor non-Abelian quasiparticles, features the so-called particle-hole Pfaffian topological order. However, the possibility that the edge modes were not thermally equilibrated left room for alternative interpretations. To avoid this ambiguity, Dutta *et al*. created an isolated edge consisting only of a 1/2 charge mode and a neutral edge mode by interfacing the fractional 5/2 state with integer quantum Hall states. The measurement of the thermal conductance of this isolated state reaffirmed the particle-hole Pfaffian character of the 5/2 state. —JS

Science, abm6571, this issue p. 1198

CANCER

Understanding prostate cancer plasticity

Cellular plasticity is the ability of cells to alter their identity and acquire new biological properties. In the context of cancer, plasticity may explain how certain tumors evolve and become resistant to therapy. Chan *et al*. investigated the molecular mechanisms underlying lineage plasticity in late-stage prostate cancer. Using several model systems, including human organoid cultures, activation of Janus kinase–signal transducer and activator of transcription (JAK/STAT) and fibroblast growth factor receptor (FGFR) signaling was found to drive prostate tumor plasticity. Pharmacological targeting of the JAK and FGFR pathways promoted lineage reprogramming in biopsy samples from castration-resistant prostate cancer patients. These findings provide

insights that may help to curb prostate cancer progression and resensitize resistant tumors to effective therapies. —PNK

Science, abn0478, this issue p. 1180

EXOPLANETS

Separating rocky from watery exoplanets

The radii of small transiting exoplanets have a bimodal distribution, usually interpreted as planets with and without a thick atmosphere. Luque and Pallé demonstrate that the planets' densities (which require measurements of both mass and radius) provide a cleaner separation between the populations than radii alone, at least for exoplanets orbiting red dwarf stars (see the Perspective by Teske). They found that one group is consistent with a purely rocky composition, whereas another group matches a model of 50% rock and 50% water. This split between rocky and water-rich planets could reflect where they formed in their planetary systems, before orbital migration. —KTS

Science, abl7164, this issue p. 1211; see also add7175, p. 1156

METASURFACES

Another twist for metasurfaces

Metasurfaces are specially designed arrays of dielectric components that transform the function of bulk optical components into thin films. Exploiting the physics of bulk states in the continuum for the highly efficient trapping of light, Zhang *et al.* demonstrate metasurfaces that operate as a source of chiral light (see the Perspective by Forbes). Using a dielectric metasurface doped with light-emitting molecules, the authors were able to produce chiral photoluminescence and lasing. This approach will be useful for the development of integrated optical devices. —ISO

Science, abq7870, this issue p. 1215; see also add5065, p. 1152

TUMOR IMMUNOLOGY

Uncovering the source of adverse events

Despite the success of immune checkpoint blockade (ICB) in cancer, the efficacy is limited by immune-related adverse events (irAEs) that require treatment cessation. Identifying peptides that induce both antitumor and irAE responses is crucial to improving ICB. Berner *et al*. developed a method for detecting these peptides, which they call DITAS (discovery of tumor-associated self-antigens). DITAS involves analyzing the shared antigens between non–small cell lung cancer (NSCLC) and lung tissue and then performing immunopeptidomics, determining which HLA these peptides bound to, predicting the $CD8^+$ T cell epitopes of these peptides, performing functional validation, and finally identifying their candidate antigens. Using this method, the authors identified the self-antigen napsin A, which induced NSCLC and lung-reactive T cells. —DAE

Sci. Immunol. **7**, eabn9644 (2022).

GUT MICROBIOME

A c-di-AMP gut check for *C. difficile*

Clostridioides difficile causes antibiotic-associated colitis. Oberkampf *et al*. report that this bacterium produced and degraded the intracellular second messenger cyclic-di-adenosine monophosphate (c-di-AMP) in response to different environmental conditions (see the Focus by Purcell). Production of c-di-AMP promoted the growth of *C. difficile* but induced sensitivity to osmotic stress. Bile salts induced c-di-AMP degradation, which was required for resistance to hyperosmotic stress and the antibiotic action of bile salts. Mutant bacteria that could not degrade c-di-AMP did not colonize the guts of antibiotic-treated mice. —AMV

Sci. Signal. **15**, eabn8171, eadd3937 (2022).

REVIEW SUMMARY

LAND USE CHANGE

Disentangling the numbers behind agriculture-driven tropical deforestation

Florence Pendrill*, Toby A. Gardner*, Patrick Meyfroidt, U. Martin Persson, Justin Adams, Tasso Azevedo, Mairon G. Bastos Lima, Matthias Baumann, Philip G. Curtis, Veronique De Sy, Rachael Garrett, Javier Godar, Elizabeth Dow Goldman, Matthew C. Hansen, Robert Heilmayr, Martin Herold, Tobias Kuemmerle, Michael J. Lathuillière, Vivian Ribeiro, Alexandra Tyukavina, Mikaela J. Weisse, Chris West

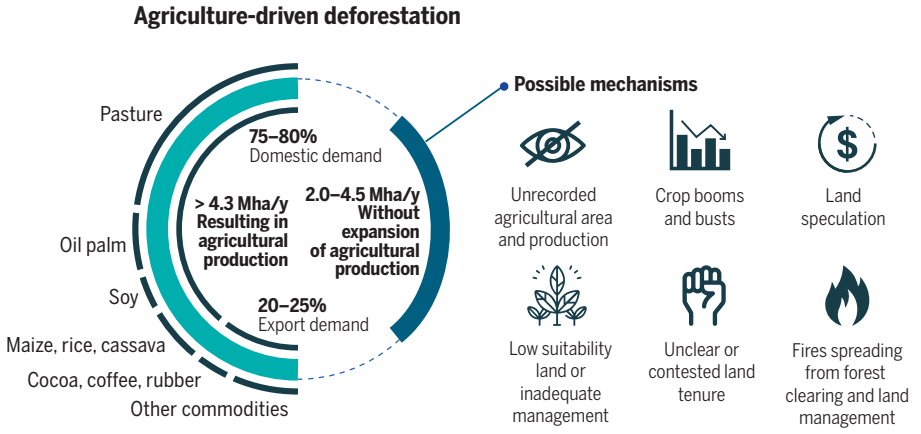
BACKGROUND: Agricultural expansion is a primary cause of tropical deforestation and therefore a key driver of greenhouse gas emissions, biodiversity loss, and the degradation of ecosystem services vital to the livelihoods of forest-dependent and rural people. However, agriculture-driven deforestation can take many forms, from the direct expansion of pastures and cropland into forests to more complex or indirect pathways. A clear understanding of the different ways in which agriculture drives deforestation is essential for designing effective policy responses. To address this need we provide a review of the literature on pantropical agriculture-driven deforestation and synthesize the best available evidence to quantify dominant agricultural land-use changes relating to deforestation. We consider the policy implications of this assessment, especially for burgeoning demand-side and supply-chain interventions seeking to address deforestation.

ADVANCES: New methods and data have advanced our understanding of deforestation and subsequent land uses. However, only a handful of studies estimate agriculture-driven deforestation across the entirety of the tropics. Although these studies agree that agriculture is the dominant land use following forest clearing, their estimates of pantropical rates of agriculture-driven deforestation during the period 2011 to 2015 vary greatly—between 4.3 and 9.6 million hectares (Mha) per year—with our synthesized estimate being 6.4 to 8.8 Mha per year. This apparent uncertainty in the amount of agriculture-driven deforestation can be disentangled by distinguishing between the different ways in which agriculture contributes to deforestation; we find that while the overwhelming majority (90 to 99%) of tropical deforestation occurs in landscapes where agriculture is the dominant driver of tree cover loss, a smaller share (45 to 65%) of deforestation is due to the expansion of active agricultural production into forests. Multiple lines of evidence show that the remainder of agriculture-driven deforestation does not result in the expansion of productive agricultural land but instead is a result of activities such as speculative clearing, land tenure issues, short-lived and abandoned agriculture, and agriculture-related fires spreading to adjacent forests.

Different land uses and commodities often interact to drive deforestation. However, pasture expansion is the most important driver by far, accounting for around half of the deforestation resulting in agricultural production across the tropics. Oil palm and soy cultivation together account for at least a fifth, and six other crops—rubber, cocoa, coffee, rice, maize, and cassava—likely account for most of the remainder, with large regional variations and higher levels of uncertainty.

OUTLOOK: This Review points to three key areas where a stronger evidence base would advance global efforts to curb agriculture-driven deforestation: First, consistent pantropical data on deforestation trends are lacking. This limits our ability to assess overall progress on reducing deforestation and account for leakage across regions. Second, with the exception of soy and oil palm the attribution of deforestation to forest risk commodities is often based on coarse-grained agricultural statistics, outdated or modeled maps, or local case studies. Third, uncertainties are greatest in dry and seasonal tropics and across the African continent in particular.

This assessment highlights that although public and private policies promoting deforestation-free international supply chains have a critical role to play, their ability to reduce deforestation on the ground is fundamentally limited. One-third to one-half of agriculture-driven deforestation does not result in actively managed agricultural land. Moreover, the majority—approximately three-quarters—of the expansion of agriculture into forests is driven by domestic demand in producer countries, especially for beef and cereals, including much of the deforestation across the African continent. These data suggest that the potential for international supply chain measures to help reduce tropical deforestation is more likely to be achieved through interventions in deforestation risk areas that focus on strengthening sustainable rural development and territorial governance. ■



Agriculture contributes to deforestation in many ways which often interact. Most tropical deforestation occurs in landscapes where agriculture is the dominant driver of forest loss. Part of this agriculture-driven deforestation results in agricultural production (left) meeting domestic and export demand for various agricultural commodities. However, agriculture-driven deforestation also occurs without expansion of managed agricultural land through several mechanisms (right), which may lead to the deforested area being abandoned or semi-abandoned. Incomplete agricultural records also explain a share of such deforestation.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: florence.pendrill@chalmers.se (F.P.); toby.gardner@sei.org (T.A.G.)
Cite this article as F. Pendrill *et al.*, *Science* **377**, eabm9267 (2022). DOI: 10.1126/science.abm9267

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm9267

REVIEW

LAND USE CHANGE

Disentangling the numbers behind agriculture-driven tropical deforestation

Florence Pendrill[1]*, Toby A. Gardner[2]*, Patrick Meyfroidt[3,4], U. Martin Persson[1], Justin Adams[5], Tasso Azevedo[6], Mairon G. Bastos Lima[2], Matthias Baumann[7], Philip G. Curtis[8], Veronique De Sy[9], Rachael Garrett[10,11], Javier Godar[2], Elizabeth Dow Goldman[12], Matthew C. Hansen[13], Robert Heilmayr[14,15], Martin Herold[16], Tobias Kuemmerle[7,17], Michael J. Lathuillière[2], Vivian Ribeiro[2], Alexandra Tyukavina[13], Mikaela J. Weisse[12], Chris West[18]

Tropical deforestation continues at alarming rates with profound impacts on ecosystems, climate, and livelihoods, prompting renewed commitments to halt its continuation. Although it is well established that agriculture is a dominant driver of deforestation, rates and mechanisms remain disputed and often lack a clear evidence base. We synthesize the best available pantropical evidence to provide clarity on how agriculture drives deforestation. Although most (90 to 99%) deforestation across the tropics 2011 to 2015 was driven by agriculture, only 45 to 65% of deforested land became productive agriculture within a few years. Therefore, ending deforestation likely requires combining measures to create deforestation-free supply chains with landscape governance interventions. We highlight key remaining evidence gaps including deforestation trends, commodity-specific land-use dynamics, and data from tropical dry forests and forests across Africa.

Deforestation continues at high rates, mainly in the tropics (*1*–*4*), and is one of the largest drivers of greenhouse gas emissions, biodiversity loss, and the degradation of ecosystem services (*5*). Although deforestation is driven by many interrelated processes (*6*), expanding agricultural land use—including cropland, pastures, and tree crops—is the primary direct cause of tropical deforestation (*7*–*9*).

[1]Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden. [2]Stockholm Environment Institute (SEI), Stockholm, Sweden. [3]Georges Lemaître Earth and Climate Research Centre, Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium. [4]Fonds de la Recherche Scientifique F.R.S.-FNRS, Brussels, Belgium. [5]Tropical Forest Alliance, World Economic Forum, Geneva, Switzerland. [6]Observatório do Clima, MapBiomas, São Paulo, Brazil. [7]Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany. [8]Juniata Analytics LLC, Denver, CO, USA. [9]Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, Wageningen, Netherlands. [10]Environmental PolicyLab, Department of Humanities, Social, and Political Sciences, ETH Zurich, Zürich, Switzerland. [11]Department of Geography and Cambridge Conservation Initiative, Cambridge University, Cambridge, UK. [12]Global Forest Watch, World Resources Institute, Washington, DC, USA. [13]Department of Geographical Sciences, University of Maryland, College Park, Maryland, USA. [14]Environmental Studies Program, University of California, Santa Barbara, Santa Barbara, California, USA. [15]Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, California, USA. [16]Helmholz GFZ Research Centre for Geosciences, Section 1.4 Remote Sensing and Geoinformatics, Telegrafenberg, Potsdam, Germany. [17]Integrated Research Institute for Transformations in Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Berlin, Germany. [18]Stockholm Environment Institute York, Department of Environment and Geography, University of York, York, UK.
*Corresponding authors. Email: florence.pendrill@chalmers.se (F.P.); toby.gardner@sei.org (T.A.G.)

Currently there is considerable attention on curbing tropical deforestation, including renewed commitments to reduce deforestation at the climate COP26 in 2021, upcoming negotiations at the COP15 for the Convention on Biological Diversity, and strengthened commitments and legislative proposals from governments (*10*–*12*), companies (*13*, *14*), and financial institutions (*15*). Emerging policies often focus on eliminating deforestation from international supply chains of agrifood commodities such as palm oil, soybeans, and beef. With the adequacy of past pledges having received damning assessments (e.g., the New York Declaration on Forests 5-year assessment in 2019), largely due to lack of funding and implementation, it is crucial that renewed investment is guided by the best available evidence on agriculture-driven deforestation. The targeting of limited resources needs to have its basis in a clear understanding of the scale of the problem, its location, and the relative importance of different drivers.

However, at present policies are being designed and evaluated against a backdrop of widespread uncertainty regarding our understanding of the links between agriculture and deforestation. The focus on agricultural supply chain policies is commonly premised on statements that agricultural expansion and production drive 80% of tropical deforestation, a number appearing in everything from policy proposals [e.g., by the EU (*10*) and the UK (*16*)], to high-profile research (e.g., *17*), and communications from NGOs and international organizations [e.g., Rainforest Alliance (*18*), Greenpeace (*19*)]. This 80% number frequently appears as fact without referencing the original source, Hosonuma *et al.* (*20*), or understanding its meaning and limitations. In 2012, the referenced study gave a much needed "first inventory of what countries identify as relevant and important drivers" (*21*). However, data sources and methods for identifying deforestation and subsequent land uses have since improved considerably (*1*, *2*, *7*, *22*–*25*). At this critical juncture of the fate of the world's tropical forests, it is essential to take stock of our current understanding of the role agriculture plays in driving deforestation and to identify key data and knowledge gaps.

We aim to provide such a synthesis to disentangle the key rates and mechanisms of agriculture-driven deforestation, organized around three central questions regarding our current understanding of (i) the rates and trends in deforestation across the tropics, (ii) the role of agriculture in driving deforestation, both in terms of the direct expansion of productive agricultural land and more broadly regarding the links between agriculture and land use dynamics (e.g., land speculation), and (iii) the relative importance of different forest risk commodities in driving deforestation and the extent to which their production is linked to international trade. We assess our ability to address these questions in different regions, clarify the inherent challenges in quantifying the role of agriculture in driving tropical deforestation, and consider the practical implications of existing knowledge for science and policy.

Agriculture and deforestation

The drivers or causes of deforestation can be examined in many ways (*26*, *27*) and there are often interactions between multiple drivers (*6*, *9*). This Review focuses on agriculture-driven deforestation, here defined broadly as deforestation for which agriculture—whether directly or indirectly—is a cause (Box 1). Notably, agriculture-driven deforestation is not limited to the direct expansion of commodity production into forests. We review recent pantropical assessments of deforestation drivers (table S1) and complement this with a literature search of national-level estimates for eleven countries with the highest deforestation rates (*28*). We harmonized datasets to the same set of 87 tropical and subtropical countries (henceforth the "tropics"), covering most of Latin America, Africa south of the Sahara, and South and Southeast Asia (*28*) (fig. S1), and focus on the time period of 2011 to 2015.

Deforestation rates and trends

Estimating deforestation rates across the tropics presents both conceptual and technical challenges. First, there is no single way to distinguish between forests and nonforests nor between deforestation and forest degradation and as a result different studies and monitoring

systems rely on different definitions (*29*–*31*). Second, although remote sensing is useful for monitoring forest changes in terms of land cover not all aspects of deforestation—including its underlying drivers—can be observed from satellites; further, technical and practical constraints result in imperfect data (e.g., negative effects of cloud cover) (*29*, *30*). Forest loss estimates therefore differ between studies (fig. S2) both because of measurement uncertainties (*32*) and because they strive to measure different things.

We define deforestation as a persistent conversion of natural forest to any other land use, such as agriculture or human settlements, or to tree plantations (Box 1). This definition aligns with the aims of many policies focused on the loss of natural forests and concomitant losses of biodiversity, carbon stocks, and other ecosystem services, and builds upon the definition presented by the Accountability Framework initiative (*33*). There is currently no pantropically consistent, spatially explicit dataset that quantifies deforestation as defined above, although Vancutsem *et al.* (*2*) comes close for tropical moist forests (*28*). Therefore, this Review combines data from different sources to derive estimates in line with that definition.

The two main global data sources on forest loss used by most in the policy and research communities are the Global Forest Change (GFC) dataset based on annual, remote-sensing–based measures of tree cover loss (TCL) (*1*) and the Food and Agriculture Organization (FAO)'s Forest Resources Assessment (FRA) which reports deforestation rates at 5- to 10-year intervals (*3*). Many recent pantropical assessments of deforestation drivers rely partly on GFC. A key challenge for assessing deforestation based on the GFC data is that although all deforestation is in principle captured by TCL, not all TCL (a land cover change) constitutes deforestation in terms of a persistent change in land use away from natural forest (*1*, *28*) (Fig. 1A). In particular, TCL includes clearings within tree plantations, severe forest degradation, and rotational cycles of shifting cultivation (*1*, *7*). The FRA uses a more restrictive definition of deforestation than the one used here in which conversion of natural forest to forestry plantation is not considered deforestation. Its usefulness for assessing deforestation drivers is limited as the data are compiled at the national level only and are collected from country reports based on a variety of methods including remote sensing and inventories (*34*).

For 2011 to 2015, GFC TCL rates averaged 10.6 million hectares (Mha) per year in the tropics while the FAO FRA 2020 estimated deforestation to be 10.7 Mha per year (Fig. 1B), despite the latter applying a more restrictive definition and primarily reporting net (not gross) deforestation for many countries. These aggregated numbers mask considerable regional differences especially for Africa where FRA deforestation is estimated at 4.4 Mha per year, although TCL amounts to 2.8 Mha per year (Fig. 2 and table S3). For some countries these differences are also considerable: for India, the FRA (gross) deforestation rate (0.67 Mha per year; based on remote sensing) (*3*) far exceeds the GFC TCL (0.10 Mha per year).

Box 1. Key terms for disentangling agriculture-driven deforestation.

Natural forest: A forest that "resembles—in terms of species composition, structure and ecological function—one that is or would be found in a given area in the absence of major human impact" (*33*). Aside from primary and intact forests, natural forest also includes regenerated (second-growth) forests and partially degraded forests, provided they fulfill the definition above (*33*). As no comprehensive pantropical map of natural forests currently exists, most studies approximate their extent.

Deforestation: A persistent conversion of natural forest to any other land use such as agriculture, human settlements, or tree plantations.

Agriculture: Agriculture includes cropland, pastures, and tree crops but not forestry (agriculture thus excludes timber, pulp, and paper).

Agriculture-driven deforestation: Deforestation for which agriculture, directly or indirectly, is a cause; this includes both deforestation resulting in agricultural production and agriculture-driven deforestation without expansion of agricultural production. Agriculture-driven deforestation does not necessarily mean that agriculture is the only or main cause of deforestation; for example, deforestation may be directly driven by the demand for timber alongside the demand for agricultural expansion (*49*, *50*, *95*); further, indirect or underlying drivers always play a role (*6*, *27*).

Deforestation resulting in agricultural production: This is deforestation that can be attributed to the expansion of land under active agricultural production systems.

Agriculture-driven deforestation without expansion of agricultural production: This is defined as deforestation occurring in landscapes where agriculture is the dominant driver of forest loss but does not result in recorded, productive, and actively managed agricultural land. This can be due to several mechanisms and is distinct from forest degradation or other TCL in the sense that the forest has been fully cleared and there are signs of other types of land use, though in practice the boundary can be hard to draw.

Additionally, the two main datasets show opposing pantropical trends between 2001 to 2010 and 2011 to 2020, with an increase from 9 to 12 Mha per year in GFC TCL (*1*), compared with a decrease from 14 to 10 Mha per year in FRA deforestation rates (*3*) (Fig. 1B). Although discrepancies in rates are expected as approaches differ in how they define forests and deforestation [for further discussion see (*1*, *2*, *32*, *35*)], the observation that the GFC TCL and FRA deforestation data report a difference in the overall trend direction is more puzzling. Uncertainties in trends arise as a result of several methodological and conceptual challenges which must be taken into account for drawing conclusions about trends in TCL or deforestation based on the GFC and FAO FRA datasets, e.g., Curtis *et al.* (*7*), Carter *et al.* (*32*), Goldman *et al.* (*36*), Pendrill *et al.* (*37*), and Nguyen and Kanemoto (*38*).

The increasing trend in GFC TCL presents two main challenges for evaluating temporal trends in deforestation: First, the GFC methodology has become more effective at detecting small and temporary forest disturbances—some of which could be more adequately characterized as forest degradation rather than deforestation—after 2011 and especially after 2015 (*39*, *40*), due to both changes in the methodology and increased quality and volume of Landsat satellite data. Caution is thus needed when trying to compare TCL trends between the pre- and post-2011 or -2015 time periods (*28*, *39*, *40*). Second, this effect is enhanced by the growing importance of forest degradation, which has increased in many parts of the tropics in recent years as a result of the combined effects of climate change, fires, forest fragmentation, and unsustainable timber extraction (*2*, *41*, *42*).

For the FRA 2020 deforestation data, "relatively few countries and territories have reliable data over the [full] period" (*43*). There has been some evidence that "countries with lower capacities in the past had the tendency to overestimate the area of forest loss" (*44*). In recent years data sources have improved for many tropical countries (*34*, *43*) potentially leading to inconsistencies with older data of lower quality. The decreasing trend in the FRA deforestation rates may thus in part result from overestimates and uncertainties in earlier years [though decelerating deforestation is also found in the preliminary (global) results from the remote sensing study accompanying the FRA 2020 (*4*)]. Overall we thus find that consistent pantropical data on deforestation trends is lacking, challenging our ability to assess if and where progress is being made.

Agriculture-driven deforestation

There are currently only a handful of pantropical estimates of the importance of agriculture in

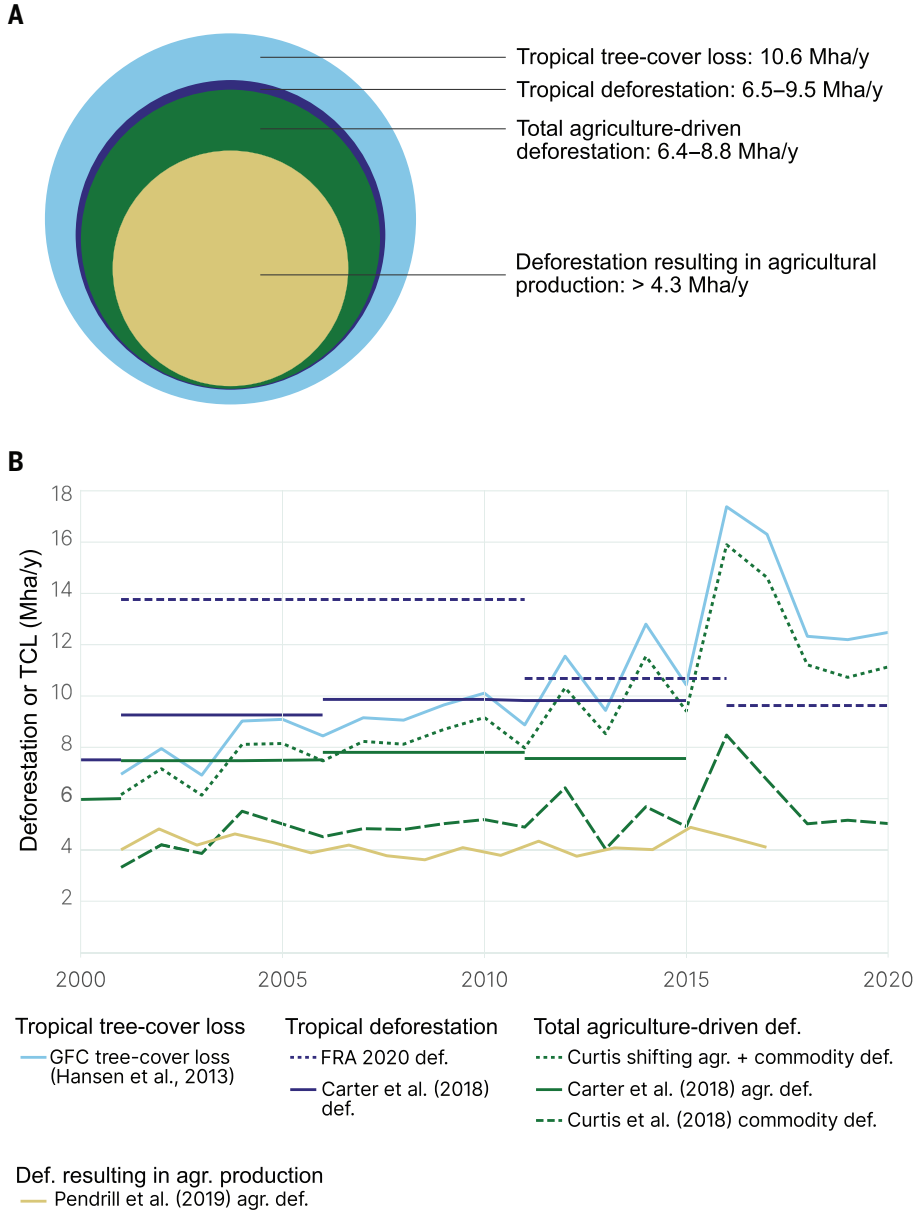


Fig. 1. TCL, deforestation, and agriculture-driven deforestation. (**A**) Conceptual diagram visualizing the concepts of tropical TCL, deforestation, agriculture-driven deforestation, and deforestation resulting in agricultural production, nested from the broadest to the narrowest concept. The area of each circle is scaled by the estimated extent although the ranges are not represented; consequently for deforestation and agriculture-driven deforestation the extent is approximated. (**B**) Studies vary considerably in their estimated extents (millions of hectares per year) and trends, reflecting uncertainties and conceptual differences. The TCL data are from GFC [updated from Hansen *et al.* (*1*)]; deforestation, from the FAO FRA (*3*) and Carter *et al.* (*32*); agriculture-driven deforestation, updated from Curtis *et al.* (*7*), Carter *et al.* (*32*), and deforestation resulting in agricultural production updated from Pendrill *et al.* (*37*). def, deforestation; agr, agriculture. In all figures, the data have been aligned to the same set of 87 (sub)tropical countries.

driving deforestation (*7*, *8*, *20*, *32*, *37*, *45*) (table S1), all of which agree that agriculture is the dominant land use following deforestation. Estimates of deforestation drivers, e.g., the relative importance of agriculture and of different commodities, are intrinsically less reliable in the most recent years, because time is needed to reveal whether the cleared land will be used for production (and if so, for what) or allowed to regenerate. Typically use of the cleared land is assessed within at least two to four years after forest clearing although the precise number of years varies between studies (~1 to <20 years) depending on method and data availability (*28*). For these reasons we focus our analysis on the period 2011 to 2015.

For that period three studies provide pantropical estimates of agriculture-driven deforestation (fig. S3). One [Carter *et al.* (*32*)] assumes a constant fraction of deforestation being agriculture-driven on the basis of pre-2010 data from other studies [De Sy *et al.* (*8*) and Hosonuma *et al.* (*20*)]. The other two—despite relying on the same GFC TCL data (*1*)—provide notably different estimates of agriculture-driven deforestation ranging from 4.3 Mha per year [Pendrill *et al.* (*37*)] up to 9.6 Mha per year [Curtis *et al.* (*7*)] (Fig. 1B and table S4). This variation arises both as a result of methodological differences and because estimates describe different aspects of deforestation and the role of agriculture therein.

By combining these two assessments [Curtis *et al.* (*7*) and Pendrill *et al.* (*37*)] with ancillary data (*28*) we estimate total agriculture-driven deforestation across the tropics to be 6.4 to 8.8 Mha per year (Fig. 1A). As detailed below, this range reflects uncertainties in how much of the TCL attributable to shifting agriculture constitutes deforestation as opposed to cyclical crop-fallow rotations. With total deforestation ranging between 6.5 and 9.5 Mha per year (table S3), this implies that most (~90 to 99%) tropical deforestation occurs in landscapes where agriculture is the dominant driver of forest loss (*28*).

The Pendrill *et al.* (*37*) data suggest a much smaller share of tropical deforestation resulting in agricultural production, in the range of ~45 to 65% of our total tropical deforestation estimate (likely at the higher end) (*28*). Pendrill *et al.* (*37*) estimate this by employing a land-balance model to attribute GFC TCL to expanding cropland and pastures. They evaluate the expansion of cropland and pastures primarily on the basis of national agricultural statistics [FAOSTAT (*46*), with subnational data for Brazil and Indonesia]. A key source of uncertainty in the Pendrill *et al.* (*37*) assessment comes from its reliance on FAOSTAT-recorded agricultural areas. The quality of these data varies considerably between countries, and data are often imputed or estimated rather than reported (Table 1) (*46*). This can lead to underestimation of the significance of agriculture as a deforestation driver for countries that are slower to (or simply do not) update their statistics and where the self-reporting by countries incompletely capture some agricultural activities (e.g., shifting cultivation). The Pendrill *et al.* (*37*) estimate of 4.3 Mha per year of deforestation resulting in agricultural production should therefore be considered a conservative estimate (*28*).

By contrast, Curtis *et al.* (*7*) assess the dominant direct drivers of TCL in 10-by-10 km grid cells using decision tree models trained on high-resolution imagery in Google Earth. Dominant drivers of GFC TCL are divided into five classes: commodity-driven deforestation

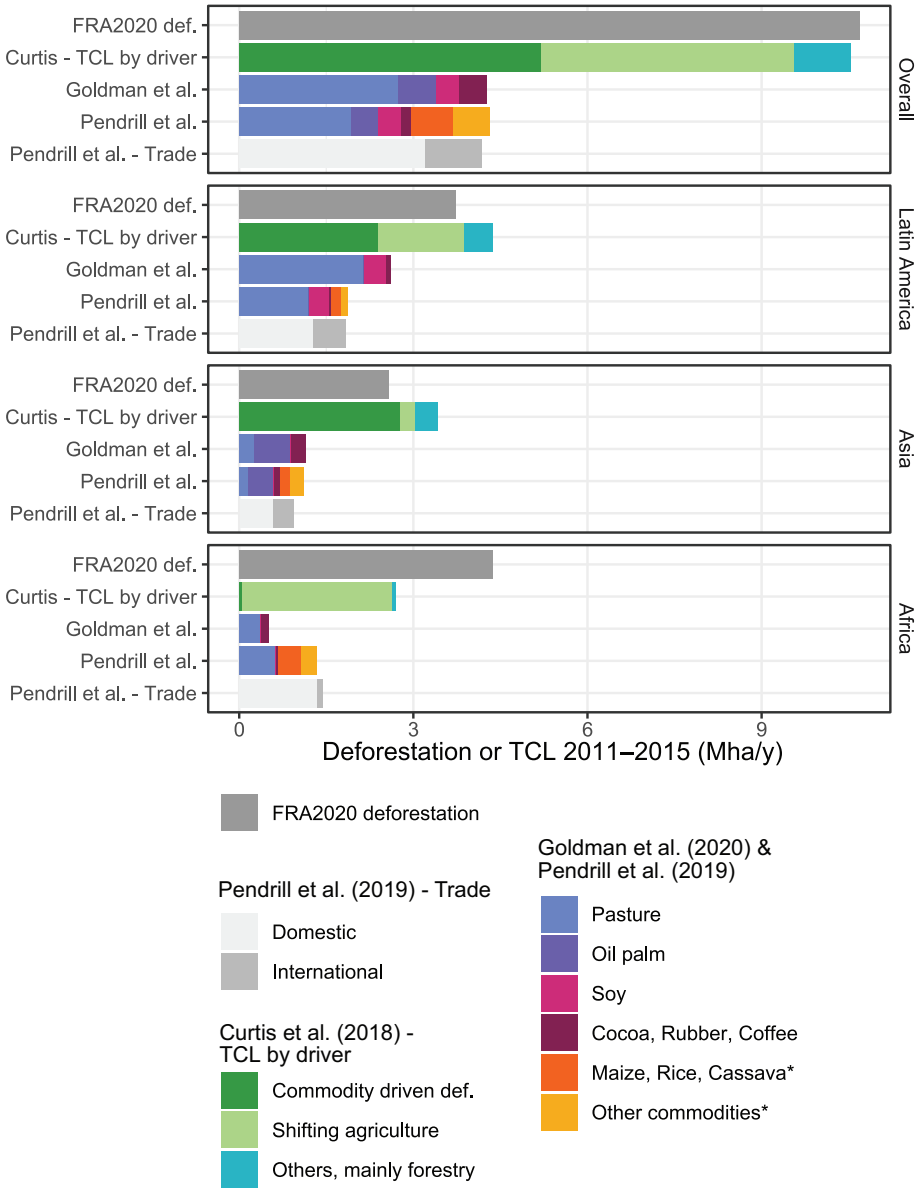


Fig. 2. Estimates of tropical deforestation and its agricultural drivers. The average extents (2011 to 2015) of TCL by driver [data from Hansen *et al.* (*1*) and Curtis *et al.* (*7*), where TCL driven by agriculture falls under shifting agriculture and commodity-driven deforestation] and of deforestation attributed to agricultural commodities [data from Goldman *et al.* (*36*) and Pendrill *et al.* (*37*)] and international trade [data from Pendrill *et al.* (*37*)]. Commodities followed by an asterisk are not quantified by Goldman *et al.* (*36*). FAO FRA (*3*) deforestation rates are included for comparison. TCL, tree cover loss; def, deforestation; agr, agriculture; prod, production.

(5.19 Mha per year; primarily for agriculture), shifting agriculture (4.37 Mha per year), forestry (0.93 Mha per year), wildfire (0.02 Mha per year), and urbanization (0.02 Mha per year).

For assessing agriculture-driven deforestation, the Curtis *et al.* (*7*) approach presents two key challenges. First, it does not fully distinguish which of the GFC TCL is deforestation. Some of the dominant drivers of TCL correspond to deforestation (i.e., commodity-driven deforestation and urbanization) but others do not (i.e., wildfires potentially resulting in regrowth). Still, the large remainder—i.e., shifting agriculture and forestry—can reflect both the expansion of these systems into natural forests (i.e., deforestation) and regular rotations in stable shifting agriculture systems, plantations, or managed forests, which do not constitute deforestation under most definitions (including the one adopted here). Second, the Curtis *et al.* (*7*) approach allocates all TCL in each grid cell to a single dominant (defined as >50%) driver of TCL for the whole time period (2001 to 2020), ignoring drivers that are not dominant. Therefore, even in the grid cells where commodity-driven deforestation or shifting agriculture is the dominant driver of TCL, not all TCL is necessarily directly driven by agricultural expansion. The Curtis *et al.* (*7*) estimate is thus a metric of deforestation occurring in landscapes where agriculture is the dominant direct driver of forest loss (rather than only deforestation resulting in agricultural production per se).

This metric deviates conceptually from our definition of agriculture-driven deforestation as remote sensing data can never unambiguously distinguish deforestation indirectly driven by agriculture from drivers that are colocated but causally uncoupled. However, drivers of deforestation often interact (*6*, *9*, *47*) and so in landscapes where most deforestation is directly due to agriculture, evidence from multiple studies suggest that agriculture typically contributes indirectly to much of the deforestation directly driven by other factors (*6*, *48*). For example, in agricultural deforestation frontiers even if logging or urbanization is the direct driver of some deforestation it is typically indirectly linked to agriculture; e.g., areas in which land is logged first but with prospects of conversion to agriculture, which may or may not materialize (*49*–*51*), or where urbanization is connected to the inflow of laborers into agriculture (*52*). The share of deforestation in pixels where Curtis *et al.* (*7*) classify agriculture as the dominant driver but which is causally disconnected from agriculture is therefore likely to be very small. Hence, we take the metric of deforestation occurring in landscapes where agriculture is the dominant direct driver of forest loss as the best available proxy for estimating agriculture-driven deforestation.

Curtis *et al.* (*7*) put deforestation occurring in landscapes where agriculture is the dominant driver in the range of 5.19 Mha per year (commodity-driven deforestation only) to 9.47 Mha per year (sum of commodity-driven deforestation and shifting agriculture) (Fig. 2). We narrowed this range down to 6.4 to 8.8 Mha per year (*28*) by excluding TCL in tree plantations (*53*) and by including deforestation in primary forests (*54*) and deforestation resulting in agricultural production [based on Pendrill *et al.* (*37*)] (fig. S4).

Our analysis suggests a large discrepancy (2.0 to 4.5 Mha per year) between the deforestation resulting in agricultural production (>4.3 Mha per year) and the overarching category of agriculture-driven deforestation (6.4 to 8.8 Mha per year) (Figs. 1A and 3). This discrepancy is present across all three continents in our country sample, totaling 1.0 to 2.0 Mha per year in Latin America, 0.0 to 1.3 Mha per year in Africa, and 1.1 to 1.2 Mha per year in Asia (Fig. 3), though uncertainties abound and part of the discrepancy is likely a result of unrecorded agricultural areas.

The discrepancy reflects the complex role of agriculture as a driver of tropical deforestation and indicates that around one-third to one-half of agriculture-driven deforestation does not result in recorded agricultural land (though it might be used for other purposes). This is consistent with regional and pantropical remote sensing studies finding large tracts of unused land following forest loss (*8*, *24*, *28*, *55*, *56*), including a pantropical estimate that 20 to 30% of agriculture-driven TCL in the period 2015 to 2019 showed some shrub or forest regrowth by 2020 (*57*).

There are several mechanisms explaining this large share of agriculture-driven deforestation without expansion of agricultural production. One such mechanism is land speculation, often linked to unclear or contested tenure. This process has been documented for several Latin American countries including the Brazilian Amazon (*58*, *59*) and Costa Rica (*60*), where expectations about future agricultural rents—fueled by planned road infrastructure improvements, uncertainties around future forest conservation policies, and the existence of large tracts of undesignated public land—lead to speculative clearing. Other social processes such as imitation [see (*61*, *62*), for example] create crop booms and potential busts (*63*). This can lead to land being cleared anticipatively but not subsequently being taken into production because of deteriorating market conditions, failed operations, or diminishing economic viability. For instance, land cleared for speculation in the Brazilian Amazon is typically put under extensive pasture where animal stocking rates are very low; these pastures are commonly degraded and abandoned within relatively short time periods (*64*–*66*). Deforestation can also be used to strengthen tenure claims where laws link land rights to clearing or use (*67*, *68*). Moreover, conflicts over land tenure often contribute to deforestation in contested forest frontiers, in excess of clearings purely for productive agriculture (*69*, *70*). The extent of land with unclear and contested tenure is not precisely quantified pantropically but is shown to be very large in some countries (*71*). Land degradation can also lead to land abandonment or maintenance of the land at very low levels of productivity, possibly because the deforested land was not suitable to begin with (*72*, *73*) or because of deforestation-driven changes in local climate (*74*), inadequate management and lack of expertise, or cultural or structural barriers (*66*, *75*).

Another mechanism through which agriculture contributes to deforestation without resulting in productive agricultural land in the near term is from fires started in agricultural lands that spread to adjacent forest areas, leading to forest degradation and in some cases complete deforestation. Almost all fires in tropical moist forests are due to human activities (*42*) including clearing forests for new agriculture and as a land management tool (e.g., weed control and nutrient mobilization) in already-cleared agricultural areas (*42*). This frequently leads to fires spreading into adjacent forest areas as documented in Brazil (*76*), the Miombo (*77*), and Indonesia (*78*).

Attributing deforestation to commodities and consumers

The evidence on pantropical rates of deforestation attributed to cropland, pasture, and associated commodity production in more recent

Table 1. Data availability for assessing deforestation resulting in agricultural production. Deforestation rates (total and for major post–forest loss land uses, in millions of hectares per year) for the 11 countries with the highest rates of deforestation in the period 2011 to 2015, and quality of the underlying driver data (cell shading). Estimates are from Pendrill *et al*. (*37*) (P), Goldman *et al*. (*36*), or other studies (O) identified in the literature review and where national-level estimates for the time period 2011 to 2015 could be extracted from the source (*28*). Superscripts indicate the citations to other studies (O). Superscript a indicates citations for cropland deforestation in Brazil (*130*) and Indonesia (*55*, *114*); superscript b for pasture deforestation in Brazil (*131*); superscript c for soybean deforestation in Brazil (*80*, *132*, *133*); superscript d for oil palm deforestation in Indonesia (*55*, *114*, *134*, *135*) and Malaysia (*136*, *137*).

	Deforestation rate	Cropland		Pasture			Soybeans			Oil palm			Rubber		Cocoa		Coffee		Maize, rice, cassava		
		P	O[a]	P	G	O[b]	P	G	O[c]	P	G	O[d]	P	G	P	G	P	G	P	P	P
Latin America																					
Brazil	1.5–2.2	0.46	0.19	0.75	1.1	0.49	0.27	0.22	0.06-0.16	0.00	0.00		0.00	0.00	0.00	0.02	0.00	0.02	0.11	0.01	0.01
Paraguay	0.36–0.38	0.11		0.14	0.14		0.08	0.02		0.00	0.00						0.00	0.00	0.01	0.01	0.00
Argentina	0.28–0.33	0.00		0.00	0.13		0.00	0.08			0.00							0.00	0.00	0.00	0.00
Bolivia	0.20–0.24	0.02		0.04	0.34		0.00	0.04			0.00				0.00	0.00	0.00	0.00	0.00	0.00	0.00
Africa:																					
DR Congo	0.37–0.84	0.36		0.02	0.01		0.00	0.00		0.01	0.00		0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.08	0.12
Angola	0.18	0.02		0.18	0.02		0.00	0.00		0.00	0.00				0.00	0.00	0.00	0.00	0.01	0.00	0.01
Madagascar	0.07–0.26	0.00		0.01	0.04		0.00	0.00		0.00	0.00				0.00	0.00	0.00	0.02	0.00	0.00	0.00
Mozambique	0.17	0.00		0.18	0.03			0.00			0.00							0.00	0.00	0.00	0.00
Asia:																					
Indonesia	1.2–1.3	0.64	0.3-0.8	0.09	0.03		0.00	0.01		0.39	0.45	0.14-0.24	0.04	0.06	0.01	0.05	0.00	0.03	0.03	0.07	0.00
Malaysia	0.25–0.26	0.07		0.00	0.01		0.00	0.00		0.05	0.16	0.08	0.01	0.05	0.00		0.00	0.00	0.00	0.01	0.00
Myanmar	0.14–0.24	0.06		0.01	0.06		0.00	0.00			0.00		0.01				0.00	0.00	0.01	0.00	0.00

Def. rate (Mha/y)
0 0.1 0.3 ∞

Data quality classification:
Recent multitemporal extent maps of high resolution (<=30 m or vector) and/or accuracy.
Recent (>2012), single year extent maps of high spatial resolution (<=30 m or vector).
Official subnational agricultural statistics (recent & multitemporal, but not spatially explicit).
Official national-level agricultural statistics (recent & multitemporal, but not spatially explicit).
Based on unofficial national-level agricultural statistics (e.g., imputed by the FAO) or on older, coarse-resolution maps.

years primarily stems from only two approaches: Pendrill *et al.* (*37*) and Goldman *et al.* (*36*). Two other studies have also quantified the role of agricultural commodity production in driving deforestation (*38*, *45*) but these primarily cover time periods before 2010 and are thus not discussed in detail here. Pendrill *et al.* (*37*) is the most comprehensive in terms of commodity coverage, with annual data on deforestation followed by pasture and 155 crops, assessed primarily at the national level. Given its lack of spatial detail this method does not unequivocally establish whether these land uses expanded directly on cleared forest land or indirectly displaced other land uses into the forest (*37*). Goldman *et al.* (*36*) attribute deforestation to commodities by overlaying GFC TCL classified as commodity-driven deforestation or shifting agriculture [from Curtis *et al.* (*7*)] with recent spatially explicit extent maps for oil palm, soy, rubber, and pasture for a subset of countries as well as older, coarse maps for pasture, cocoa, and coffee. The coarse estimates are far more uncertain (than those based on recent maps) for two main reasons. First, all TCL classified as dominated by commodity-driven deforestation or shifting agriculture is assumed to constitute deforestation resulting in agricultural production, which risks over-allocating TCL as deforestation assigned to commodities. Second, it assumes that the relative shares of commodity area and thus share of deforestation in each grid cell have remained stable since the year 2000 for pasture and 2010 for crops. This is unlikely to hold especially in rapidly changing deforestation frontiers.

Fig. 3. The ways in which agriculture contributes to deforestation differ between regions. Agriculture-driven deforestation [based on Curtis *et al.* (*7*)] includes deforestation resulting in agricultural production [based on Pendrill *et al.* (*37*)] as well as agriculture-driven deforestation without expansion of agricultural production, which can occur through several potential mechanisms. Incomplete records of agricultural area and production might also explain a share of that deforestation, which should thus be attributed to certain land uses and commodities if monitoring systems improve. Deforestation resulting in agricultural production can in turn be attributed further to certain land uses and commodities [based on Pendrill *et al.* (*37*) and Goldman *et al.* (*36*)], and to export or domestic demand [based on Pendrill *et al.* (*37*)].

It is well established that cattle pasture expansion is the single most important deforestation driver by far, alone accounting for around half of the deforestation resulting in agricultural production (*36*, *37*). Still, the two available pantropical datasets differ considerably in the estimated extent of deforestation attributed to the expansion of pastures [1.9 compared with 2.7 Mha per year with the lower value from Pendrill *et al.* (*37*) and the higher from Goldman *et al.* (*36*)]. Most of the deforestation due to the expansion of pastures is found in South America (~1.2 and 2.1 Mha per year) (Fig. 2), particularly in Brazil. This region has robust data on pasture-driven deforestation at the national or biome level (table S5). Attributing deforestation to pasture is especially challenging (*28*) because of its complex dynamics with other drivers (e.g., land speculation and crops) (*58*, *79*–*81*); additionally, pastures can be difficult to distinguish from other land cover types based on remote sensing as they may appear spectrally similar to cropland or natural vegetation (*82*, *83*) and because pastures and their definitions vary considerably (*84*, *85*).

Following pasture the next most important land uses are oil palm and soy cultivation, together accounting for at least a fifth of the deforestation resulting in agricultural production (*36*, *37*). Their importance is reflected in the large number of country or biome-wide assessments of these crops (table S5) (*28*). Deforestation attributed to these crops is highly concentrated regionally, notably in South America for soy and in Southeast Asia for oil palm (Fig. 2 and table S6), particularly in Indonesia. Pantropical estimates are also the most reliable for these two crops (Table 1) though precise estimates can still differ from and between national-specific studies [e.g., Indonesia (*28*)], underscoring the value of having multiple data sources.

The cultivation of rubber, cocoa, coffee, rice, maize, and cassava account for most of the

remaining deforestation resulting in agricultural production (*28*, *36*, *37*). However, the evidence is currently lacking to confidently estimate their significance or changes in their significance over time (*37*) and country-level assessments are largely lacking (table S5). For these crops the data are limited or of poor quality (Table 1) and both pantropical approaches rely heavily on agricultural statistics. Statistical records are unreliable for cocoa and coffee cultivation (*86*) with further uncertainties as these crops can be shade grown, in which case their expansion into natural forest can be difficult to detect using remote sensing (*87*–*89*). Further, they are also often grown together with other crops in agroforestry systems (*87*–*89*). Records for staple crops are frequently based on estimates and may underestimate harvested areas in subsistence or smallholder contexts as a result of minimum harvested area criteria in records (*90*).

Many of the crops discussed above are important export crops—including soybeans, palm oil, rubber, coffee, and cacao—and international trade has been identified as a key driver of deforestation since the 2000s (*89*, *91*–*93*). Three pantropical studies assess deforestation associated with trade in commodities: Nguyen and Kanemoto (*38*), Cuypers *et al*. (*45*), and Pendrill *et al*. (*37*). The first two are not discussed further as their deforestation data are primarily for the pre-2010 time period.

The role of international demand in driving deforestation differs depending on how far downstream the analysis extends in regards to international supply chains (*94*). A physical trade model—which traces deforestation embodied in raw or lightly processed agricultural commodities—suggests that 20 to 25% of all deforestation resulting in agricultural production is linked to exports (*37*) (fig. S5). This average, however, hides substantial variation across countries, regions, and commodities (fig. S6): soybeans, palm oil, and cash crops (e.g., rubber, coffee, and cocoa) are primarily destined for export markets whereas beef and cereals are typically consumed domestically. An economic, multiregional input-output model, which traces deforestation all the way to final consumption, raises the share of commodity-driven deforestation linked to international demand to ~35% (*37*) (fig. S5). Thus despite the remaining limitations and uncertainties in data and current trade models there is convincing evidence that domestic demand remains a primary underlying driver of deforestation resulting in agricultural production.

Although the numbers presented here provide a big-picture indication of the most important forest risk commodities, commodities often interact in driving deforestation. Deforestation can also be followed by several successive agricultural land uses (*28*). For example, in South America soy expansion in one location has been linked to pasture expansion in others (*79*, *81*), and timber harvesting is often a precursor to deforestation, e.g., in case of oil palm expansion in Indonesia (*49*, *95*). Such concurrent and interacting drivers of forest degradation and deforestation are poorly evaluated in continental-scale assessments, which can lead to an overly simplified focus on addressing drivers in isolation (*47*, *96*). Additionally, data are largely lacking on the legality of deforestation and production (*97*), whether the actors involved are small- or large-holders, and whether they are producing for subsistence or marketed demand (*98*–*100*).

Moreover, we have not assessed nonagricultural deforestation drivers. Logging and demand for wood products (e.g., timber and pulp), charcoal, and fuelwood—alongside agricultural expansion—are key direct drivers of deforestation and even more so of degradation (*6*, *55*, *101*, *102*). Although deforestation resulting from the expansion of tree plantations is estimated by Goldman *et al*. (*36*) and Pendrill *et al*. (*37*) (0.1 and 0.8 Mha per year, respectively, with the former only covering eight countries), deforestation due to logging and timber extraction that sometimes occurs in conjunction with and facilitates agriculture expansion (*49*, *50*, *95*) is not comprehensively quantified at the pantropical level.

Urbanization, mining, and energy infrastructure such as hydropower dams are relatively minor direct drivers of deforestation from a pantropical perspective, together amounting to just 2% of land use following forest loss across the (sub)tropics between 1990 and 2000 (*8*), although they can be important direct drivers locally; e.g., gold mining is a dominant direct cause of deforestation in Guyana (*103*) and in Madre de Dios in Peru (*104*). Further, the indirect impacts of these drivers can be considerable (*71*, *105*–*107*). A study of the Brazilian Amazon found that deforestation indirectly induced by mining was 12 times the extent of that of the direct deforestation occurring within mining concessions (*108*).

Improving the evidence base

Our findings point to three key data gaps in our understanding of tropical deforestation and its links to agriculture. Overcoming these gaps can considerably strengthen the evidence base to help accelerate global efforts to curb agriculture-driven deforestation—both in the design of policy responses and in evaluating their effectiveness.

First, the lack of consistent pantropical data on deforestation still hampers our ability to assess overall deforestation trends and thus the net impacts of interventions to reduce deforestation while accounting for leakage across regions and biomes (*109*–*111*). Improvements in deforestation data are needed in three main areas: (i) to encompass both dry and wet tropics, (ii) to provide estimates of deforestation that go beyond TCL and satisfy the commonly held definition of a persistent conversion of natural forest to any other land use, and (iii) to ensure that estimates are consistent across regions and over time. Data on deforestation trends could be improved in several ways to help meet these requirements, including improvement of contextual data on tree plantations and shifting agriculture to systematically filter out temporary TCL from the GFC data (*1*), or by expanding the Vancutsem *et al*. (*2*) approach to the dry tropics, for example. Furthermore, deforestation area metrics alone are a crude proxy for multiple social-ecological impacts, which vary significantly between geographical locations (*30*). Improved quantification of these impacts is an ongoing need.

Second, to improve our understanding of the relationships between agricultural drivers and forest loss and to inform both territorial and supply chain measures directed at specific commodities, a concerted effort is needed to improve the coverage, quality, and frequency of data on pastures and crops that are replacing forests for all regions where substantial deforestation occurs. In contrast to deforestation data, data on drivers need not be pantropical as commodity-specific deforestation frontiers are typically concentrated in specific regions and require responses tailored to their context (*111*). Regional-level datasets that can cover most of a given commodity, e.g., soy across South America and oil palm in Southeast Asia, play a key role as they are built on regional knowledge and thus are typically not just more accurate but also more region- and policy-relevant, e.g., in terms of land use and management characterization (*112*). However, currently only oil palm (*113*) and soy (*25*) are mapped for most production areas in the tropics (*36*). The attribution of deforestation and conversion to most forest risk commodities, especially outside of Brazil and Indonesia, therefore relies on agricultural statistics at a very coarse—often national—scale, on local case studies, or on single-year, modeled maps that are often outdated, potentially leading to misattribution. Despite the fact that pastureland is by far the most prominent driver of deforestation, our understanding of pasture extent is particularly poor as large-scale assessments outside of South America rely on (often unofficial) agricultural statistics or on a global pasture map for the year 2000 (*28*).

Important recent advances in land use mapping include multiple biome-scale initiatives such as MapBiomas (*114*), sample-based monitoring tools such as CollectEarth (*115*), and efforts to combine wall-to-wall satellite monitoring and sample-based approaches, including building confidence in temporal trends in deforestation (*4*, *23*–*25*, *116*, *117*). Future advances should include improving the collection

and organization of subnational agricultural statistics and further leveraging advances in remote-sensing data and methods (*8*, *22*).

Third, there is an urgent need to invest in spatially and temporally explicit assessments of agriculture-driven deforestation tailored to the dry tropics and to deforestation frontiers in Africa, with a focused effort to better characterize deforestation in smallholder shifting agriculture [e.g., (*100*)]. Uncertainties around the nature, extent, and drivers of deforestation linked to agriculture are unevenly distributed as the quality of the data used and the performance of the methods vary between countries and biomes (*1*, *2*, *7*, *32*, *36*) (Table 1). Overall, our understanding of agriculture-driven deforestation is systematically poorer in dry forests and wooded savannas and across the African continent, in contrast with Latin America and the humid tropics. There are several reasons for this: First, there is a general neglect of land-use change research in Africa (*9*), where, additionally, the capacity of agencies to compile data on agricultural production is particularly limited (*118*, *119*). Our literature search found comparatively fewer studies on recent agriculture-driven deforestation in Africa (n = 6), compared with Latin America (n = 27) and Asia (n = 26) (table S5). Tropical dry forests are also less researched than wet forests (*116*, *120*). Second, remote-sensing mapping of forests and agricultural land cover and their changes is generally more difficult in heterogeneous landscapes, e.g., where tree cover and canopy structure vary, and where smallholder and shifting agriculture results in small, irregularly shaped, temporally dynamic patches of cultivated land interspersed with natural vegetation (*1*, *121*, *122*). These challenges are exacerbated by difficulties in discriminating vegetation types for intermediate levels of tree cover, such as in savannas, shrublands, and sparsely forested woodlands, which are more prevalent on the African continent (*30*, *77*, *116*).

This disparity in our understanding of the dry and seasonal tropics compared with the wet tropics (Table 1) is particularly striking given that about one-third of all tropical dry forests and woodlands are in active deforestation frontiers (*56*). Further emphasis on deforestation in the dry and seasonal tropics would also challenge the disproportionate prioritization of international conservation funding towards moist forest biomes (*123*).

Conclusions

The synthesis of current data on agriculture-driven deforestation provided here challenges conventional wisdom and has profound implications for policy. The central insight from our Review is the distinction—and discrepancy—between agriculture-driven deforestation and deforestation resulting in active agricultural production. Although as much as 90 to 99% of deforestation occurs in landscapes where agriculture is the main driver of TCL, only 45 to 65% of deforestation can be attributed to the expansion of actively managed cropland, pasture, or tree crops. The implications of this discrepancy are wide-ranging for efforts to curb deforestation and to mitigate climate change. The most recent global carbon budget indicates a stagnation or decline in global emissions from land use change, due most notably to reduced tropical cropland expansion (*124*). However, that assessment does not account for forest degradation or the large share of deforestation not resulting in agricultural production identified here. The discrepancy also highlights two essential conclusions that can shape more effective policy responses to deforestation.

First, although public and private policies promoting deforestation-free international supply chains have a key role to play (*96*, *125*), their direct effectiveness in reducing deforestation is fundamentally limited given that (i) international demand represents only a quarter of total deforestation resulting in agricultural production, and (ii) one third to one half of agriculture-driven deforestation does not result in productive agricultural land. Additionally, most supply-chain interventions to date have been focused on direct sourcing and are restricted in their ability to address products associated with deforestation that enter supply chains through intermediaries (*126*). International supply chain interventions can, in principle, help address some of the indirect ways agriculture drives deforestation (e.g., by discouraging speculative clearings) (*127*). However, tackling deforestation linked with domestic demand as well as the underlying drivers of agriculture-driven deforestation more broadly, such as land-tenure insecurities and conflicts, likely requires also broader land governance and rural development interventions (*125*, *128*). Tenure reform, land zoning, regulatory reform and enforcement, and extension services supporting farmers, all have an important role to play in slowing agriculture-driven deforestation (*125*, *128*, *129*). Many of these approaches would likely benefit from closer partnerships between demand and supply side actors and the scaling up of deforestation-free supply chains to deforestation-free regions and sectors. There is an urgent need to identify and leverage the mechanisms by which demand-side supply-chain policies, including zero-deforestation commitments, can go beyond their immediate impacts and help motivate and catalyze broader changes in territorial governance. This remains a key research frontier.

Second, to effectively reduce deforestation, interventions need to address the systemic interdependencies between the expansion of different commodities, requiring a much stronger focus on more comprehensive, landscape-level approaches. The most prominent example of this is pasture expansion, which is tightly linked to soy expansion and land speculation across Latin America. An excessive focus on individual commodities, which characterize many current policy initiatives, risks undermining the potential to avoid widespread leakage and deliver positive reductions in deforestation on the ground.

The unprecedented focus on forest conservation and nature-based climate solutions in the aftermath of the UNFCCC COP 26 and heading into the UN Biodiversity COP 15 provides a critical moment to ensure that urgent efforts to tackle deforestation are guided and evaluated by an evidence base fit for purpose.

REFERENCES AND NOTES

1. M. C. Hansen *et al*., High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013). doi: 10.1126/science.1244693; pmid: 24233722
2. C. Vancutsem, *et al*., Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Sci. Adv.* **7**, eabe1603 (2021). doi: 10.1126/sciadv.abe1603
3. FAO, Global Forest Resources Assessment (FAO, 2020); https://fra-data.fao.org/.
4. FAO, FRA 2020 Remote Sensing Survey (FAO, 2022); https://www.fao.org/forest-resources-assessment/remote-sensing/fra-2020-remote-sensing-survey/en/.
5. IPBES, "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services" E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo, Eds. (IPBES, 2019); https://doi.org/10.5281/zenodo.3831673.
6. H. J. Geist, E. F. Lambin, Proximate causes and underlying driving forces of tropical deforestation. *Bioscience* **52**, 143–150 (2002). doi: 10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2
7. P. G. Curtis, C. M. Slay, N. L. Harris, A. Tyukavina, M. C. Hansen, Classifying drivers of global forest loss. *Science* **361**, 1108–1111 (2018). doi: 10.1126/science.aau3445; pmid: 30213911
8. V. De Sy *et al*., Tropical deforestation drivers and associated carbon emission factors derived from remote sensing data. *Environ. Res. Lett.* **14**, 094022 (2019). doi: 10.1088/1748-9326/ab3dc6
9. J. Busch, K. Ferretti-Gallon, What drives deforestation and what stops it? A meta-analysis. *Rev. Environ. Econ. Policy* **11**, 3–23 (2017). doi: 10.1093/reep/rew013
10. European Commission, Deforestation and forest degradation – reducing the impact of products placed on the EU market (European Commission, 2021); https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12137-Deforestation-and-forest-degradation-reducing-the-impact-of-products-placed-on-the-EU-market.
11. FOREST Act of 2021, S.2950, 117th Cong. (2021).
12. UK Environment Act 2021, Schedule 17 Use of forest risk commodities in commercial activity (2021); https://www.legislation.gov.uk/ukpga/2021/30/schedule/17/enacted#schedule-17.
13. J. Luciano *et al*., Agricultural commodity companies corporate statement of purpose (UN Climate Change Conference, 2021); https://ukcop26.org/agricultural-commodity-companies-corporate-statement-of-purpose/.
14. "Nurturing transparency: The path to forest positive" (Consumer Goods Forum 2021 Annual Report, 2021); https://www.theconsumergoodsforum.com/wp-content/uploads/2021/09/CGF-FPC-Annual-Report-2021.pdf.
15. ACTIAM, Andra AP-fonden (AP2), Australian Ethical Investment, Aviva PLC, AXA Group, Boston Common Asset Management, Church Commissioners for England, Church of England Pensions Board, Comgest, Caisse de Prévoyance de L'État de Genève (CPEG), Domini Impact Investments LLC, East Capital Group, Fidelity International, Generation Investment Management, Grupo Bancolombia, Impax Asset Management PLC, International Business of Federated

Hermes, JGP Asset Management, Legal and General Investment Management (LGIM), LGPS Central Limited, Lombard Odier, Menhaden PLC, NEI Investments, NN Investment Partners, Robeco, Schroders, SCOR SE, Skandia, Sparebank 1 Forsikring, Storebrand Asset Management, Sumitomo Mitsui Trust Asset Management, TCI Fund Management Ltd, The Local Authority Pension, Fund Forum, Financial sector commitment letter on eliminating commodity-driven deforestation (2021); https://racetozero.unfccc.int/wp-content/uploads/2021/11/DFF-Commitment-Letter-.pdf.
16. Government response to the recommendations of the Global Resource Initiative (2020); https://www.gov.uk/government/publications/global-resource-initiative-taskforce-government-response/government-response-to-the-recommendations-of-the-global-resource-initiative.
17. M. Crippa *et al*., Food systems are responsible for a third of global anthropogenic GHG emissions. *Nat. Food* **2**, 198–209 (2021). doi: 10.1038/s43016-021-00225-9
18. Rainforest Alliance, Our mission to protect the world's forests (2019); https://www.rainforest-alliance.org/insights/our-mission-to-protect-the-worlds-forests/.
19. Greenpeace, Agribusiness & deforestation, (2021); https://www.greenpeace.org/usa/forests/issues/agribusiness/.
20. N. Hosonuma *et al*., An assessment of deforestation and forest degradation drivers in developing countries. *Environ. Res. Lett.* **7**, 044009 (2012). doi: 10.1088/1748-9326/7/4/044009
21. G. Kissinger, M. Herold, V. De Sy, "Drivers of deforestation and forest degradation: a synthesis report for REDD+ policymakers" (Lexeme Consulting, 2012); https://www.cifor.org/knowledge/publication/5167/.
22. R. N. Masolele *et al*., Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series. *Remote Sens. Environ.* **264**, 112600 (2021). doi: 10.1016/j.rse.2021.112600
23. P. Potapov *et al*., Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat. Food* **3**, 19–28 (2022). doi: 10.1038/s43016-021-00429-z
24. V. Zalles *et al*., Rapid expansion of human impact on natural land in South America since 1985. *Sci. Adv.* **7**, eabg1620 (2021). doi: 10.1038/s43016-021-00429-z
25. X.-P. Song *et al*., Massive soybean expansion in South America since 2000 and implications for conservation. *Nat. Sustain.* **2021**, 784–792 (2021). doi: 10.1038/s41893-021-00729-z; pmid: 34377840
26. A. M. Hersperger, M.-P. Gennaio, P. H. Verburg, M. Bürgi, Linking land change with driving forces and actors: Four conceptual models. *Ecol. Soc.* **15**, art1 (2010). doi: 10.5751/ES-03562-150401
27. P. Meyfroidt, Approaches and terminology for causal analysis in land systems science. *J. Land Use Sci.* **11**, 501–522 (2016). doi: 10.1080/1747423X.2015.1117530
28. Materials and methods are available as supplementary materials.
29. R. L. Chazdon *et al*., When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. *Ambio* **45**, 538–550 (2016). doi: 10.1007/s13280-016-0772-y; pmid: 26961011
30. J. O. Sexton *et al*., Conservation policy and the measurement of forests. *Nat. Clim. Chang.* **6**, 192–196 (2016). doi: 10.1038/nclimate2816
31. A. I. Fernández-Montes de Oca *et al*., An integrated framework for harmonizing definitions of deforestation. *Environ. Sci. Policy* **115**, 71–78 (2021). doi: 10.1016/j.envsci.2020.10.007
32. S. Carter *et al*., Agriculture-driven deforestation in the tropics from 1990–2015: Emissions, trends and uncertainties. *Environ. Res. Lett.* **13**, 014002 (2018). doi: 10.1088/1748-9326/aa9ea4
33. Accountability Framework, Terms and definitions (Accountability Framework initiative, 2020); https://accountability-framework.org/wp-content/uploads/2020/03/Definitions-Mar2020.pdf.
34. M. K. Nesha *et al*., An assessment of data sources, data quality and changes in national forest monitoring capacities in the Global Forest Resources Assessment 2005-2020. *Environ. Res. Lett.* **16**, 054029 (2021). doi: 10.1088/1748-9326/abd81b
35. A. Tyukavina *et al*., Congo Basin forest loss dominated by increasing smallholder clearing. *Sci. Adv.* **4**, eaat2993 (2018). doi: 10.1126/sciadv.aat2993; pmid: 30417092
36. E. D. Goldman, M. Weisse, N. Harris, M. Schneider, Estimating the role of seven commodities in agriculture-linked deforestation: Oil palm, soy, cattle, wood fiber, cocoa, coffee, and rubber (World Resources Institute, 2020); https://www.wri.org/research/estimating-role-seven-commodities-agriculture-linked-deforestation-oil-palm-soy-cattle.
37. F. Pendrill *et al*., Agricultural and forestry trade drives large share of tropical deforestation emissions. *Glob. Environ. Change* **56**, 1–10 (2019). doi: 10.1016/j.gloenvcha.2019.03.002
38. N. T. Hoang, K. Kanemoto, Mapping the deforestation footprint of nations reveals growing threat to tropical forests. *Nat. Ecol. Evol.* **5**, 845–853 (2021). doi: 10.1038/s41559-021-01417-z; pmid: 33782576
39. Global Forest Watch, Assessing trends in tree cover loss over 20 years of data (Global Forest Watch, 2021); https://www.globalforestwatch.org/blog/data-and-research/tree-cover-loss-satellite-data-trend-analysis/.
40. M. C. Hansen *et al*., University of Maryland, High-Resolution Global Maps of 21st-Century Forest Cover Change, v1.7, Earth Engine Partners (2021); http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html.
41. D. van Wees *et al*., The role of fire in global forest loss dynamics. *Glob. Change Biol.* **27**, 2377–2391 (2021). doi: 10.1111/gcb.15591; pmid: 33694227
42. P. M. Brando *et al*., Droughts, wildfires, and forest carbon cycling: A pantropical synthesis. *Annu. Rev. Earth Planet. Sci.* **47**, 555–581 (2019). doi: 10.1146/annurev-earth-082517-010235
43. FAO, "Global Forest Resources Assessment 2020: Main report" (FAO, 2020); https://www.fao.org/3/ca9825en/ca9825en.pdf.
44. E. Romijn *et al*., Assessing change in national forest monitoring capacities of 99 tropical countries. *For. Ecol. Manage.* **352**, 109–123 (2015). doi: 10.1016/j.foreco.2015.06.003
45. D. Cuypers *et al*., "The impact of EU consumption on deforestation: Comprehensive analysis of the impact of EU consumption on deforestation" (European Commission Technical Report 2013-063, 2013); https://ec.europa.eu/environment/forests/pdf/1.%20Report%20analysis%20of%20impact.pdf.
46. FAO, FAOSTAT database (FAO, 2017); http://www.fao.org/faostat/en/#home.
47. P. Meyfroidt *et al*., Middle-range theories of land system change. *Glob. Environ. Change* **53**, 52–67 (2018). doi: 10.1016/j.gloenvcha.2018.08.006
48. IPBES, "The IPBES assessment report on land degradation and restoration" L. Montanarella, R. Scholes, A. Brainich, Eds. (Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2018); https://ipbes.net/assessment-reports/ldr.
49. S. D. Tarigan, S. Sunarti, S. Widyaliza, Expansion of oil palm plantations and forest cover changes in Bungo and Merangin Districts, Jambi Province, Indonesia. *Procedia Environ. Sci.* **24**, 199–205 (2015). doi: 10.1016/j.proenv.2015.03.026
50. IUFRO, "Illegal logging and related timber trade – dimensions, drivers, impacts and responses" D. Kleinschmit, S. Mansourian, C. Wildburger, A. Purret, Eds. (IUFRO World Series No. 35, 2016); https://www.iufro.org/fileadmin/material/publications/iufro-series/ws35/ws35-high-res.pdf
51. P. Richards, What drives indirect land use change? How Brazil's agriculture sector influences frontier deforestation. *Ann. Assoc. Am. Geogr.* **105**, 1026–1040 (2015). doi: 10.1080/00045608.2015.1060924; pmid: 26985080
52. P. Richards, L. VanWey, Where deforestation leads to urbanization: How resource extraction is leading to urban growth in the Brazilian Amazon. *Ann. Assoc. Am. Geogr.* **105**, 806–823 (2015). doi: 10.1080/00045608.2015.1052337; pmid: 26985079
53. N. Harris, E. Goldman, S. Gibbes, Spatial Database of Planted Trees (SDPT) Version 1.0 (GFW, 2019); www.globalforestwatch.org.
54. S. Turubanova, P. V. Potapov, A. Tyukavina, M. C. Hansen, Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. *Environ. Res. Lett.* **13**, 074028 (2018). doi: 10.1088/1748-9326/aacd1c
55. K. G. Austin, A. Schwantes, Y. Gu, P. S. Kasibhatla, What causes deforestation in Indonesia? *Environ. Res. Lett.* **14**, 024007 (2019). doi: 10.1088/1748-9326/aaf6db
56. A. Buchadas, M. Baumann, P. Meyfroidt, T. Kuemmerle, Uncovering major types of deforestation frontiers across the world's tropical dry woodlands. *Nat. Sustain.* **5**, 619–627 (2022). doi: 10.1038/s41893-022-00886-9
57. Y. Feng *et al*., Doubling of annual forest carbon loss over the tropics during the early twenty-first century. *Nat. Sustain.* **5**, 444–451 (2022). doi: 10.1038/s41893-022-00854-3
58. J. Miranda, J. Börner, M. Kalkuhl, B. Soares-Filho, Land speculation and conservation policy leakage in Brazil. *Environ. Res. Lett.* **14**, 045006 (2019). doi: 10.1088/1748-9326/ab003a
59. C. Azevedo-Ramos, P. Moutinho, No man's land in the Brazilian Amazon: Could undesignated public forests slow Amazon deforestation? *Land Use Policy* **73**, 125–127 (2018). doi: 10.1016/j.landusepol.2018.01.005
60. P. C. Roebeling, E. M. T. Hendrix, Land speculation and interest rate subsidies as a cause of deforestation: The role of cattle ranching in Costa Rica. *Land Use Policy* **27**, 489–496 (2010). doi: 10.1016/j.landusepol.2009.07.002
61. V. Junquera, P. Meyfroidt, Z. Sun, P. Latthachack, A. Grêt-Regamey, From global drivers to local land-use change: Understanding the northern Laos rubber boom. *Environ. Sci. Policy* **109**, 103–115 (2020). doi: 10.1016/j.envsci.2020.04.013
62. Y. le Polain de Waroux, Capital has no homeland: The formation of transnational producer cohorts in South America's commodity frontiers. *Geoforum* **105**, 131–144 (2019). doi: 10.1016/j.geoforum.2019.05.016
63. V. Junquera, A. Grêt-Regamey, Crop booms at the forest frontier: Triggers, reinforcing dynamics, and the diffusion of knowledge and norms. *Glob. Environ. Change* **57**, 101929 (2019). doi: 10.1016/j.gloenvcha.2019.101929
64. B. B. N. Strassburg *et al*., When enough should be enough: Improving the use of current agricultural lands could meet production demands and spare natural habitats in Brazil. *Glob. Environ. Change* **28**, 84–97 (2014). doi: 10.1016/j.gloenvcha.2014.06.001
65. S. B. Hecht, The logic of livestock and deforestation in Amazonia: Considering land markets, value of ancillaries, the larger macroeconomic context, and individual economic strategies. *Bioscience* **43**, 687–695 (1993). doi: 10.2307/1312340
66. R. D. Garrett *et al*., Explaining the persistence of low income and environmentally degrading land uses in the Brazilian Amazon. *Ecol. Soc.* **22**, art27 (2017). doi: 10.5751/ES-09364-220327
67. A. Angelsen, Policies for reduced deforestation and their impact on agricultural production. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19639–19644 (2010). doi: 10.1073/pnas.0912014107; pmid: 20643935
68. B. Brito, P. Barreto, A. Brandão Jr., S. Baima, P. H. Gomes, Stimulus for land grabbing and deforestation in the Brazilian Amazon. *Environ. Res. Lett.* **14**, 064018 (2019). doi: 10.1088/1748-9326/ab1e24
69. G. M. Thaler, C. A. M. Anandi, Shifting cultivation, contentious land change and forest governance: The politics of swidden in East Kalimantan. *J. Peasant Stud.* **44**, 1066–1087 (2017). doi: 10.1080/03066150.2016.1243531
70. S. Aldrich, R. Walker, C. Simmons, M. Caldas, S. Perz, Contentious land change in the Amazon's arc of deforestation. *Ann. Assoc. Am. Geogr.* **102**, 103–128 (2012). doi: 10.1080/00045608.2011.620501
71. P. Meyfroidt *et al*., Ten facts about land systems for sustainability. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2109217118 (2022). doi: 10.1073/pnas.2109217118; pmid: 35131937
72. J. E. Laue, E. Y. Arima, Spatially explicit models of land abandonment in the Amazon. *J. Land Use Sci.* **11**, 48–75 (2016). doi: 10.1080/1747423X.2014.993341
73. S. A. Spera *et al*., Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. *Environ. Res. Lett.* **9**, 064010 (2014). doi: 10.1088/1748-9326/9/6/064010
74. Y. Malhi *et al*., Climate change, deforestation, and the fate of the Amazon. *Science* **319**, 169–172 (2008). doi: 10.1126/science.1146961; pmid: 18048654
75. O. Cortner *et al*., Perceptions of integrated crop-livestock systems for sustainable intensification in the Brazilian Amazon. *Land Use Policy* **82**, 841–853 (2019). doi: 10.1016/j.landusepol.2019.01.006
76. J. Barlow, E. Berenguer, R. Carmenta, F. França, Clarifying Amazonia's burning crisis. *Glob. Change Biol.* **26**, 319–321 (2020). doi: 10.1111/gcb.14872; pmid: 31729092
77. N. S. Ribeiro, Y. Katerere, P. W. Chirwa, I. M. Grundy, *Miombo Woodlands in a changing environment: Securing the resilience and sustainability of people and woodlands*. (Springer, 2020).
78. M. E. Cattau *et al*., Sources of anthropogenic fire ignitions on the peat-swamp landscape in Kalimantan, Indonesia.

Glob. Environ. Change **39**, 205–219 (2016). doi: 10.1016/j.gloenvcha.2016.05.005
79. P. D. Richards, R. T. Walker, E. Y. Arima, Spatially complex land change: The Indirect effect of Brazil's agricultural sector on land use in Amazonia. *Glob. Environ. Change* **29**, 1–9 (2014). doi: 10.1016/j.gloenvcha.2014.06.011; pmid: 25492993
80. H. K. Gibbs *et al*., Brazil's soy moratorium. *Science* **347**, 377–378 (2015). doi: 10.1126/science.aaa0181; pmid: 25613879
81. N. I. Gasparri, Y. le Polain de Waroux, The coupling of South American soybean and cattle production frontiers: New challenges for conservation policy and land change science. *Conserv. Lett.* **8**, 290–298 (2015). doi: 10.1111/conl.12121
82. M. M. Caldas, D. Goodin, S. Sherwood, J. M. Campos Krauer, S. M. Wisely, Land-cover change in the Paraguayan Chaco: 2000–2011. *J. Land Use Sci.* **10**, 1–18 (2015). doi: 10.1080/1747423X.2013.807314
83. H. Müller, P. Rufin, P. Griffiths, A. J. Barros Siqueira, P. Hostert, Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sens. Environ.* **156**, 490–499 (2015). doi: 10.1016/j.rse.2014.10.014
84. K.-H. Erb *et al*., Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* **553**, 73–76 (2018). doi: 10.1038/nature25138; pmid: 29258288
85. J. Oliveira *et al*., Choosing pasture maps: An assessment of pasture land classification definitions and a case study of Brazil. *Int. J. Appl. Earth Obs. Geoinf.* **93**, 102205 (2020). doi: 10.1016/j.jag.2020.102205
86. FAO, FAOSTAT methodology - Crops primary (FAO, 2018); https://fenixservices.fao.org/faostat/static/documents/QC/QC_methodology_e.pdf.
87. D. A. Hunt *et al*., Review of remote sensing methods to map coffee production systems. *Remote Sens.* **12**, 2041 (2020). doi: 10.3390/rs12122041
88. W. Niether, J. Jacobi, W. J. Blaser, C. Andres, L. Armengot, Cocoa agroforestry systems versus monocultures: A multi-dimensional meta-analysis. *Environ. Res. Lett.* **15**, 104085 (2020). doi: 10.1088/1748-9326/abb053
89. E. M. Ordway, G. P. Asner, E. F. Lambin, Deforestation risk due to commodity crop expansion in sub-Saharan Africa. *Environ. Res. Lett.* **12**, 044015 (2017). doi: 10.1088/1748-9326/aa6509
90. FAO, Crops statistics - Concepts, definitions and classifications (FAO, 2021); https://www.fao.org/economic/the-statistics-division-ess/methodology/methodology-systems/crops-statistics-concepts-definitions-and-classifications/en/.
91. R. S. DeFries, T. Rudel, M. Uriarte, M. Hansen, Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nat. Geosci.* **3**, 178–181 (2010). doi: 10.1038/ngeo756
92. A. Leblois, O. Damette, J. Wolfersberger, What has driven deforestation in developing countries since the 2000s? Evidence from new remote-sensing data. *World Dev.* **92**, 82–102 (2017). doi: 10.1016/j.worlddev.2016.11.012
93. W. R. Faria, A. N. Almeida, Relationship between openness to trade and deforestation: Empirical evidence from the Brazilian Amazon. *Ecol. Econ.* **121**, 85–97 (2016). doi: 10.1016/j.ecolecon.2015.11.014
94. K. Hubacek, K. Feng, Comparing apples and oranges: Some confusion about using and interpreting physical trade matrices versus multi-regional input–output analysis. *Land Use Policy* **50**, 194–201 (2016). doi: 10.1016/j.landusepol.2015.09.022
95. D. L. A. Gaveau *et al*., Reconciling forest conservation and logging in Indonesian Borneo. *PLOS ONE* **8**, e69887 (2013). doi: 10.1371/journal.pone.0069887; pmid: 23967062
96. R. D. Garrett *et al*., Criteria for effective zero-deforestation commitments. *Glob. Environ. Change* **54**, 135–147 (2019). doi: 10.1016/j.gloenvcha.2018.11.003
97. C. Dummett, A. Blundell, K. Canby, M. Wolosin, E. Bodnar, "Illicit harvest, complicit goods: The state of illegal deforestation for agriculture" (Forest Trends, 2021); https://www.forest-trends.org/wp-content/uploads/2021/05/Illicit-Harvest-Complicit-Goods_rev.pdf.
98. L. H. Samberg, J. S. Gerber, N. Ramankutty, M. Herrero, P. C. West, Subnational distribution of average farm size and smallholder contributions to global food production. *Environ. Res. Lett.* **11**, 124010 (2016). doi: 10.1088/1748-9326/11/12/124010
99. A. Ravikumar, R. R. Sears, P. Cronkleton, M. Menton, M. Pérez-Ojeda del Arco, Is small-scale agriculture really the main driver of deforestation in the Peruvian Amazon? Moving beyond the prevailing narrative. *Conserv. Lett.* **10**, 170–177 (2016). doi: 10.1111/conl.12264
100. A. Heinimann *et al*., A global view of shifting cultivation: Recent, current, and future extent. *PLOS ONE* **12**, e0184479 (2017). doi: 10.1371/journal.pone.0184479; pmid: 28886132
101. T. R. H. Pearson, S. Brown, F. M. Casarim, Carbon emissions from tropical forest degradation caused by logging. *Environ. Res. Lett.* **9**, 034017 (2014). doi: 10.1088/1748-9326/9/3/034017
102. E. N. Chidumayo, D. J. Gumbo, The environmental impacts of charcoal production in tropical ecosystems of the world: A synthesis. *Energy Sustain. Dev.* **17**, 86–94 (2013). doi: 10.1016/j.esd.2012.07.004
103. Guyana Forestry Commission, "Guyana REDD+ Monitoring Reporting & Verification System (MRVS)" (Guyana Forestry Commission, year 7 summary report, 2018); https://www.forestry.gov.gy/wp-content/uploads/2018/11/MRVS-Summary-Report-Year-7_November-2018_Final.pdf.
104. J. Caballero Espejo *et al*., Deforestation and forest degradation due to gold mining in the Peruvian Amazon: A 34-year perspective. *Remote Sens.* **10**, 1903 (2018). doi: 10.3390/rs10121903
105. A. Bebbington *et al*., Priorities for governing large-scale infrastructure in the tropics. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21829–21833 (2020). doi: 10.1073/pnas.2015636117; pmid: 32848071
106. J. A. Oldekop *et al*., Forest-linked livelihoods in a globalized world. *Nat. Plants* **6**, 1400–1407 (2020). doi: 10.1038/s41477-020-00814-9; pmid: 33257859
107. J. van Vliet, Direct and indirect loss of natural area from urban expansion. *Nat. Sustain.* **2**, 755–763 (2019). doi: 10.1038/s41893-019-0340-0
108. L. J. Sonter *et al*., Mining drives extensive deforestation in the Brazilian Amazon. *Nat. Commun.* **8**, 1013 (2017). doi: 10.1038/s41467-017-00557-w; pmid: 29044104
109. P. Meyfroidt *et al*., Focus on leakage and spillovers: Informing land-use governance in a tele-coupled world. *Environ. Res. Lett.* **15**, 090202 (2020). doi: 10.1088/1748-9326/ab7397
110. M. G. Bastos Lima, U. M. Persson, P. Meyfroidt, Leakage and boosting effects in environmental governance: A framework for analysis. *Environ. Res. Lett.* **14**, 105006 (2019). doi: 10.1088/1748-9326/ab4551
111. J. Börner, D. Schulz, S. Wunder, A. Pfaff, The effectiveness of forest conservation policies and programs. *Annu. Rev. Resour. Econ.* **12**, 45–64 (2020). doi: 10.1146/annurev-resource-110119-025703
112. M. G. Tulbure, P. Hostert, T. Kuemmerle, M. Broich, Regional matters: On the usefulness of regional land-cover datasets in times of global change. *Remote Sens. Ecol. Conserv.* **8**, 272–283 (2021). doi: 10.1002/rse2.248
113. A. Descals *et al*., High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth Syst. Sci. Data* **13**, 1211–1231 (2021). doi: 10.5194/essd-13-1211-2021
114. C. M. SouzaJr *et al*., Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sens.* **12**, 2735 (2020). doi: 10.3390/rs12172735
115. A. Bey *et al*., Collect Earth: Land use and land cover assessment through augmented visual interpretation. *Remote Sens.* **8**, 807 (2016). doi: 10.3390/rs8100807
116. J.-F. Bastin *et al*., The extent of forest in dryland biomes. *Science* **356**, 635–638 (2017). doi: 10.1126/science.aam6527; pmid: 28495750
117. U.S. Department of Agriculture's Forest Service (USFS), Spatial Informatics Group (SIG), World Agroforestry Centre (ICRAF), SERVIR-Mekong, "Commodity-driven forest loss: A study of Southeast Asia" (2021); https://servir.adpc.net/publications/commodity-driven-forest-loss-a-study-of-southeast-asia.
118. World Bank, "Global strategy to improve agricultural and rural statistics" (Open Knowledge Repository, Report No. 56719-GLB, 2010); https://openknowledge.worldbank.org/handle/10986/12402.
119. N. Ramankutty, A. T. Evan, C. Monfreda, J. A. Foley, Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochem. Cycles* **22**, GB1003 (2008). doi: 10.1029/2007GB002952
120. J. M. Schröder, L. P. Ávila Rodríguez, S. Günter, Research trends: Tropical dry forests: The neglected research agenda? *For. Policy Econ.* **122**, 102333 (2021). doi: 10.1016/j.forpol.2020.102333
121. A. Pérez-Hoyos, F. Rembold, H. Kerdiles, J. Gallego, Comparison of global land cover datasets for cropland monitoring. *Remote Sens.* **9**, 1118 (2017). doi: 10.3390/rs9111118
122. P. Rufin, A. Bey, M. Picoli, P. Meyfroidt, Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data. *Int. J. Appl. Earth Obs. Geoinf.* **112**, 102937 (2022). doi: 10.1016/j.jag.2022.102937
123. S. Qin *et al*., The geography of international conservation interest in South American deforestation frontiers. *Conserv. Lett.* **15**, e12859 (2022). doi: 10.1111/conl.12859
124. P. Friedlingstein *et al*., Global Carbon Budget 2021. *Earth Syst. Sci. Data* **14**, 1917–2005 (2022). doi: 10.5194/essd-14-1917-2022
125. E. F. Lambin *et al*., The role of supply-chain initiatives in reducing deforestation. *Nat. Clim. Chang.* **8**, 109–116 (2018). doi: 10.1038/s41558-017-0061-1
126. E. K. H. J. Zu Ermgassen *et al*., Addressing indirect sourcing in zero deforestation commodity supply chains. *Sci. Adv.* **8**, eabn3132 (2022). doi: 10.1126/sciadv.abn3132; pmid: 35486729
127. R. Heilmayr, L. L. Rausch, J. Munger, H. K. Gibbs, Brazil's Amazon Soy Moratorium reduced deforestation. *Nat. Food* **1**, 801–810 (2020). doi: 10.1038/s43016-020-00194-5
128. M. G. Bastos Lima, U. M. Persson, Commodity-centric landscape governance as a double-edged sword: The case of soy and the Cerrado Working Group in Brazil. *Frontiers in Forests and Global Change* **3**, 27 (2020). doi: 10.3389/ffgc.2020.00027
129. A. E. Latawiec *et al*., Improving land management in Brazil: A perspective from producers. *Agric. Ecosyst. Environ.* **240**, 276–286 (2017). doi: 10.1016/j.agee.2017.01.043
130. V. Zalles *et al*., Near doubling of Brazil's intensive row crop area since 2000. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 428–435 (2019). doi: 10.1073/pnas.1810301115; pmid: 30559198
131. E. K. H. J. Zu Ermgassen *et al*., The origin, supply chain, and deforestation risk of Brazil's beef exports. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31770–31779 (2020). doi: 10.1073/pnas.2003270117; pmid: 33262283
132. L. L. Rausch *et al*., Soy expansion in Brazil's Cerrado. *Conserv. Lett.* **12**, e12671 (2019). doi: 10.1111/conl.12671
133. E. K. H. J. zu Ermgassen *et al*., Using supply chain data to monitor zero deforestation commitments: An assessment of progress in the Brazilian soy sector. *Environ. Res. Lett.* **15**, 035003 (2020). doi: 10.1088/1748-9326/ab6497
134. D. L. A. Gaveau *et al*., Slowing deforestation in Indonesia follows declining oil palm expansion and lower oil prices. *PLOS ONE* **17**, e0266178 (2022). doi: 10.1371/journal.pone.0266178; pmid: 35349594
135. P. Noojipady *et al*., Managing fire risk during drought: The influence of certification and El Niño on fire-driven forest conversion for oil palm in Southeast Asia. *Earth Syst. Dyn.* **8**, 749–771 (2017). doi: 10.5194/esd-8-749-2017
136. O. Hamdan, K. A. Rahman, M. Samsudin, Quantifying rate of deforestation and CO2 emission in Peninsular Malaysia using Palsar imageries. *IOP Conf. Ser. Earth Environ. Sci.* **37**, 012028 (2016). doi: 10.1088/1755-1315/37/1/012028
137. D. L. A. Gaveau *et al*., Rapid conversions and avoided deforestation: Examining four decades of industrial plantation expansion in Borneo. *Sci. Rep.* **6**, 32017 (2016). doi: 10.1038/srep32017; pmid: 27605501

ACKNOWLEDGMENTS

This research contributes to the Global Land Programme glp.earth. We thank B. Loureiro and A. Sinclair for support with the figure design and J. J. Benedict for Indonesia data assistance. We also thank four anonymous reviewers for their helpful input. All errors are our own. **Funding:** This work was supported by the following: Formas grant 213:2014-1181 (to F.P. and U.M.P.); Formas grant 2016-00351 within the LEAKAGE project (to T.A.G., J.G., and U.M.P.); European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program Grant agreement 677140 MIDLAND (to P.M.); Tropical Forest Alliance (to E.D.G. and M.W.); Norway's International Climate and Forest Initiative under the Global Forest Watch Achieving Sustainability and Scaling Impact grant (to E.D.G. and M.W.); European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program Grant agreement 101001239 SYSTEMSHIFT (to T.K.); and the UK Research and Innovation

(UKRI) Global Challenges Research Fund Trade, Development and the Environmental Hub project ES/S008160/1 (to C.W.). **Author contributions:** T.A.G., J.A., P.M., F.P., and U.M.P. conceptualized the study. F.P. (lead), T.A.G., P.M., and U.M.P. (support) designed the study and the methodology, performed the analysis, and wrote the initial manuscript. All coauthors reviewed and edited the manuscript. F.P. curated the data with support from U.M.P., V.D.S., E.D.G., and M.J.W. on the pantropical data, and support from R.H., R.G., U.M.P., and V.R. on the regional and national data.

Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** All data used in this study are available from their original sources (see SM), and we express our gratitude to everyone for making their data available. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm9267
Materials and Methods
Figs. S1 to S6
Tables S1 to S7
References (*138*–*222*)

Submitted 29 October 2021; accepted 8 August 2022
10.1126/science.abm9267

Science
JOURNALS AAAS

Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLE SUMMARY

MOLECULAR BIOLOGY

Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT

Haruhiko Ehara†, Tomoya Kujirai†, Mikako Shirouzu, Hitoshi Kurumizaka*, Shun-ichi Sekine*

INTRODUCTION: In the eukaryotic cell nucleus, genomic DNA is stored as chromatin, comprising multiple nucleosomes carrying various genetic and epigenetic information. Gene transcription by RNA polymerase II (RNAPII) intrinsically affects nucleosome structures because it requires temporary unfolding of the nucleosomes to read the DNA sequence. However, RNAPII transcribes genes while maintaining the nucleosome structures, suggesting the existence of a transcription-coupled mechanism to restore the nucleosomes. However, this mechanism has remained elusive.

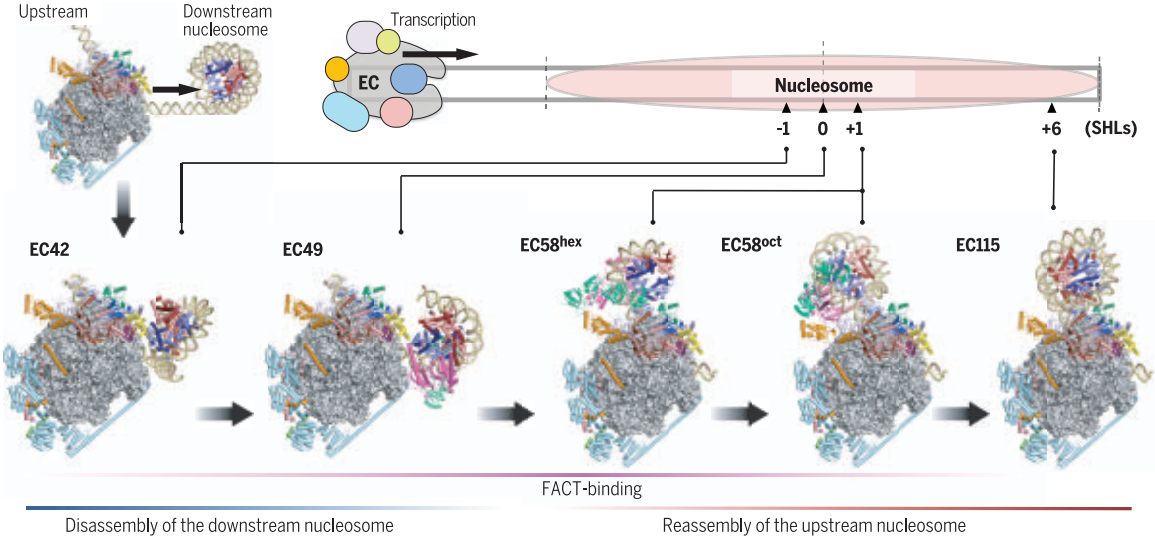
RATIONALE: We designed nucleosomal DNA templates so that the RNAPII elongation complex (EC) would stall at the 42-, 49-, 58-, and 115-bp positions from the nucleosome entry. When EC stalls at these positions, its leading edge is near super helical locations (SHLs) –1, 0, +1, and +6, respectively, of the nucleosome. Transcription was conducted in the presence of the transcription elongation factors Spn1, Spt6, Spt4/5, Elf1, Paf1C, and TFIIS and the histone chaperone FACT. The ECs formed at these positions were analyzed by cryo–electron microscopy single-particle analysis.

RESULTS: We obtained six nucleosome-transcribing EC structures: EC42, EC49, EC49B, $EC58^{hex}$, $EC58^{oct}$, and EC115, where the numbers denote the DNA positions where EC stalled. The ECs contain the RNAPII-associated Spn1, Spt6, Spt4/5, Elf1, and Paf1C proteins, which constitute the EC downstream and/or upstream edge. The structures represent serial snapshots of the EC passage through the nucleosome. In EC42, the EC leading edge resides near SHL(–1) within the downstream nucleosome, in which an ~60-bp DNA segment is removed from the histone octamer surface. One of the H2A-H2B dimers is exposed and bound with the C-terminal tail of the Spt16 subunit of FACT. In EC49, the EC leading edge is just before the nucleosomal dyad [SHL(0)], and an ~70-bp DNA segment is removed from the histone octamer. As one of the H3-H4 dimers is exposed, the main body of FACT engages with the histones primarily through Spt16. An ~30-bp DNA segment is also removed from the distal end of the nucleosome, and thus only a third of the histone octamer surface is covered by DNA. This FACT-histone complex probably represents the state before its detachment from the DNA and its subsequent transfer. In EC49B, the nucleosome is shifted downstream by ~17 bp, which might have originated by the downstream transfer of the FACT-histone intermediate. By contrast, $EC58^{hex}$ and $EC58^{oct}$, in which the EC leading edge has overrun the dyad of the original nucleosome, reveal a FACT-histone complex transferred upstream of the EC. In $EC58^{hex}$, a FACT-histone hexamer [H2A-H2B-$(H3\text{-}H4)_2$] complex is deposited onto the emerging DNA at the EC DNA exit site, and the resultant hexasome is wrapped by an ~40 bp DNA segment. In $EC58^{oct}$, the remaining H2A-H2B dimer is deposited onto the histone hexamer to form the histone octamer (octasome), which is wrapped by an ~75-bp DNA segment. Finally, EC115 reveals a near-complete nucleosome on the EC upstream side, with the histone octamer covered by an ~120-bp DNA segment. At the rim of the EC DNA exit, the domains of Spt4, Spt5, Spt6, Leo1, and Rtf1 form a "cradle" that flexibly adapts to the subnucleosome intermediates in $EC58^{hex}$, $EC58^{oct}$, and EC115, supporting the sequential nucleosome reassembly process upstream of the EC.

CONCLUSION: The obtained structures visualize key steps of nucleosome traversal by the EC accompanied by the downstream nucleosome disassembly, followed by its reassembly upstream of the EC with the aid of FACT. When the EC passes through the nucleosomal dyad, the downstream-to-upstream transfer of the nucleosomal histones occurs. These views explain the mechanism by which the nucleosome structures are maintained during transcription. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: shunichi.sekine@riken.jp (S.S.); kurumizaka@iqb.u-tokyo.ac.jp (H.K.)
†These authors contributed equally to this work.
Cite this article as H. Ehara *et al.*, *Science* **377**, eabp9466 (2022). DOI: 10.1126/science.abp9466

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abp9466



Transcription over a nucleosome mediated by the EC and FACT. Cryo–electron microscopy structures of the nucleosome-transcribing ECs: EC42, EC49, $EC58^{hex}$, $EC58^{oct}$, and EC115. EC49B is omitted in this figure. The EC contains the transcription elongation factors Spn1, Spt6, Spt4/5, Elf1, and Paf1C. These structures show snapshots of the EC progression on DNA mediating downstream nucleosome disassembly, followed by reassembly upstream of the EC with the aid of FACT.

RESEARCH ARTICLE

MOLECULAR BIOLOGY

Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT

Haruhiko Ehara[1]†, Tomoya Kujirai[1,2]†, Mikako Shirouzu[1], Hitoshi Kurumizaka[1,2]*, Shun-ichi Sekine[1]*

During gene transcription, RNA polymerase II (RNAPII) traverses nucleosomes in chromatin, but the mechanism has remained elusive. Using cryo–electron microscopy, we obtained structures of the RNAPII elongation complex (EC) passing through a nucleosome in the presence of the transcription elongation factors Spt6, Spn1, Elf1, Spt4/5, and Paf1C and the histone chaperone FACT (facilitates chromatin transcription). The structures show snapshots of EC progression on DNA mediating downstream nucleosome disassembly, followed by its reassembly upstream of the EC, which is facilitated by FACT. FACT dynamically adapts to successively occurring subnucleosome intermediates, forming an interface with the EC. Spt6, Spt4/5, and Paf1C form a "cradle" at the EC DNA-exit site and support the upstream nucleosome reassembly. These structures explain the mechanism by which the EC traverses nucleosomes while maintaining the chromatin structure and epigenetic information.

In the eukaryotic cell nucleus, genomic DNA is stored as highly organized, dense structures called chromatin. The basic unit of chromatin is the nucleosome, a disk-like assembly of an ~147-bp DNA segment wrapped around a histone octamer consisting of two copies each of the core histones H2A, H2B, H3, and H4. Arrays of nucleosomes with various genetic and epigenetic information define the chromatin structure and state, determining gene transcription patterns for the regulation of cell growth, embryonic development, and carcinogenesis. Paradoxically, gene transcription is a process that intrinsically affects the chromatin structure and epigenetic information because DNA-dependent RNA polymerase (RNAP) must temporarily disassemble nucleosomes when it passes through chromatin to transcribe the DNA sequence. However, RNAP transcribes genes while maintaining nucleosomes, suggesting the existence of a transcription-coupled mechanism to reassemble and preserve nucleosomes in the wake of the polymerase (*1*–*3*).

During gene transcription, RNAPII is associated with many protein factors, including transcription elongation factors and histone chaperones, which facilitate transcription elongation in chromatin (*1*, *2*, *4*–*9*). FACT (facilitates chromatin transcription), a histone chaperone consisting of two protein subunits (Spt16 and Pob3 in yeast; Spt16 and SSRP1 in humans), is implicated in chromatin structure establishment and maintenance (*10*, *11*). FACT is associated with nucleosomes in transcribed chromatin regions and facilitates the passage of RNAPII through the nucleosome by mediating both the disassembly of a nucleosome downstream of RNAPII and its reassembly upstream of RNAPII (*11*–*16*). This process requires multiple transcription elongation factors accompanying RNAPII. The transcription elongation factor Spt6 not only enhances RNAPII processivity (*17*–*19*) but also serves as an RNAPII-bound histone chaperone involved in nucleosome maintenance during RNAPII passage (*15*, *18*, *20*–*22*). Other elongation factors, including Spn1 (IWS1 in humans), Spt4/5 (DSIF in humans), Elf1 (ELOF1 in humans), the Paf1 complex (Paf1C, consisting of the Ctr9, Paf1, Leo1, Cdc73, and Rtf1 subunits), and TFIIS, genetically or physically interact with Spt6 and/or FACT and are implicated in chromatin transcription and integrity maintenance (*21*, *23*–*31*). Using cryo–electron microscopy (cryo-EM), in the present study, we analyzed the structures of nucleosome-transcribing RNAPII elongation complexes (ECs) obtained in the presence of these factors to gain structural insights into how the transcription machinery traverses a nucleosome by mediating its disassembly and reassembly.

[1]RIKEN Center for Biosystems Dynamics Research, Tsurumi-ku, Yokohama 230-0045, Japan. [2]Laboratory of Chromatin Structure and Function, Institute for Quantitative Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan.
*Corresponding author. Email: shunichi.sekine@riken.jp (S.S.); kurumizaka@iqb.u-tokyo.ac.jp (H.K.)
†These authors contributed equally to this work.

Results

Cryo-EM analysis

To obtain structural snapshots of RNAPII traversing a nucleosome, we designed three nucleosomal DNA templates so that RNAPII would stall at the 42-, 49-, 58-, and 115-bp positions from the nucleosome entry (Fig. 1, A and B, and methods summary). When RNAPII is stalled at these positions, its leading edge is near super helical locations (SHLs) −1, 0, +1, and +6, respectively, of the nucleosome. Transcription was conducted in the presence of the transcription elongation factors Spn1, Spt6, Spt4/5, Elf1, Paf1C, and TFIIS and the histone chaperone FACT, which facilitate the nucleosome transcription (Fig. 1A and figs. S1A and S2). The RNAPII EC stalled at the designed positions (Fig. 1C and fig. S1B). These complexes were analyzed by cryo-EM, and six EC-nucleosome structures were obtained (EC42, EC49, EC49B, EC58^{hex}, EC58^{oct}, and EC115) (Fig. 1D, figs. S3 to S14, and tables S1 and S2), as described below.

Structure of the EC

The EC contains Spn1, Spt6, Spt4/5, Elf1, and Paf1C, which cover almost the entire surface of RNAPII to form a large molecular factory (Fig. 2 and movie S1). Spn1, Elf1, and part of Paf1C (the trestle helix of the Ctr9 subunit) surround the downstream DNA, becoming components of the EC downstream edge (Fig. 2A, left). Spt4, Spt5 (the KOW1 domain), Spt6, and Paf1C (the Plus3 domain of the Rtf1 subunit and the C-terminal helix of the Leo1 subunit) surround the upstream DNA, forming the EC upstream edge (Fig. 2A, middle). Elf1 and Spt5 (the NGN domain) seal the RNAPII main channel. These interactions form an ~40-bp-long tunnel that holds the transcription bubble and parts of the downstream and upstream DNAs. Spt6 and Spt5 (the KOW4 and KOW5 domains) are integrated within the RNA exit tunnel, which surrounds approximately eight nucleotides of the exiting RNA (Fig. 2A, right). Paf1C covers the remaining surface of RNAPII except for the rim of the secondary channel, which is the TFIIS-binding site.

The EC structure reveals several unprecedented features (Fig. 2 and fig. S6). Spn1 is located at the tip of the RNAPII clamp, interacting with the Spt6 N-terminal helices and the Spt5 NGN and KOW2 domains (Fig. 2, B and C). As for Paf1C, the β-barrel domain formed between the Paf1 and Leo1 subunits binds to a site that overlaps the binding site of the homologous domain of TFIIF (Fig. 2D and fig. S15). The Leo1 C-terminal extension forms an α helix and interacts with the major groove of the upstream DNA (Fig. 2D). The Ras-like domain of the Cdc73 subunit is located near the root of the RNAPII stalk, contacting the Spt6 tSH2 domain (Fig. 2E). The C-terminal part of the Rtf1 subunit forms α helices to bind the Ctr9 subunit. More details are described in the supplementary text.

Downstream nucleosome disassembly

EC42 with the downstream nucleosome partially associated with FACT

In the EC42 structure, the EC leading edge resides near SHL(−1) within the downstream nucleosome, in which an ~60-bp segment of DNA is peeled from the histone octamer and the DNA-binding surfaces of the proximal

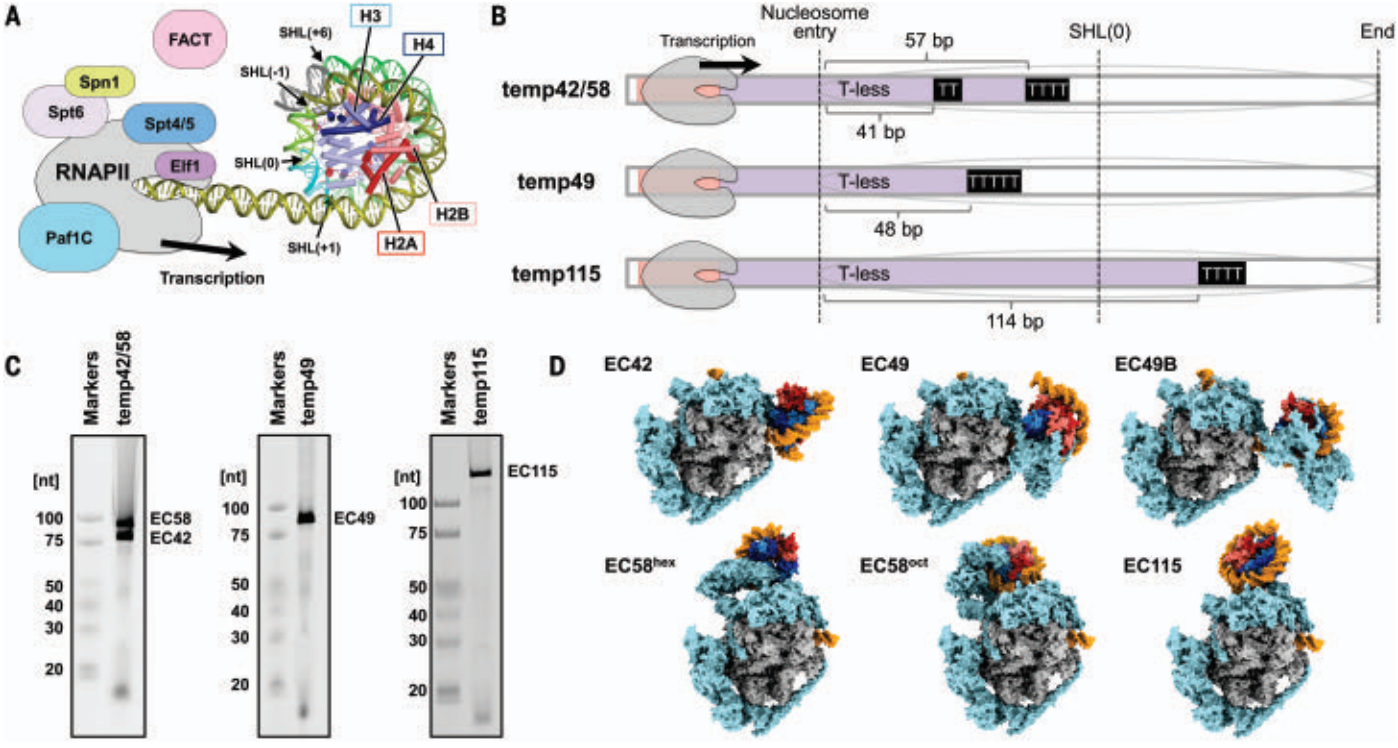


Fig. 1. Cryo-EM analysis of nucleosome transcription reactions. (**A**) Experimental setup. (**B**) Transcription templates. (**C**) Urea polyacrylamide gel electrophoresis analysis of the elongated RNAs contained in the samples analyzed by cryo-EM. (**D**) Overall structures of EC-nucleosome complexes shown in surface representations with ECs in the same orientation. RNAPII and DNA are colored gray and orange, respectively. Histones are colored as in (A). Remaining components are shown in light blue.

H2A-H2B and H3-H4 dimers are exposed (Fig. 3, A and B, and fig. S16). Herein, "proximal" and "distal" histones refer to the nucleosomal histones proximal and distal, respectively, to the transcription start position. Spn1, Spt6 (the N-terminal α helices), Elf1, and Paf1C (the Ctr9 trestle helix) cooperate with RNAPII to form the EC downstream edge, facing the partially unwrapped nucleosome (Fig. 3A, left; fig. S17; and movies S2 and S3). The nucleosome orientation relative to the EC is similar to that in the previous structure of the RNAPII-nucleosome complex containing Elf1 and Spt4/5 [SHL(−1)$_{+1B}$] (*32*), but the DNA is peeled slightly more around SHL(−1), and the DNA interaction with the H3 α2 helix is lost in EC42. Spn1 and the Spt6 N-terminal helices (residues 231 to 258) intervene between RNAPII and the nucleosomal histone octamer, and the Spn1 core domain directly contacts the exposed proximal H3-H4 dimer (Fig. 3B). Spn1 and Spt6 have flexible tails extending from the Spn1 core and the Spt6 N-terminal helices, respectively. Although these tails are not visible in the density map, they could also interact with the histones or the downstream DNA. These views are consistent with previous reports that Spn1 and Spt6 interact with histones and nucleosomes (*20*, *33*, *34*).

The Spt16 subunit of FACT comprises the N-terminal domain (NTD), dimerization domain (DD), middle domain (MD), and C-terminal domain (CTD), and the Pob3 subunit comprises the NTD/DD, MD, and CTD (*11*)(Fig. 3C). An extra density was observed on the exposed DNA-binding surface of the proximal H2A-H2B (Figs. 3B and 4A and fig. S18A), and this was interpreted as the CTD tail of the FACT Spt16 subunit, as in other FACT-nucleosome complex structures (*35*, *36*). The tail density is followed by extra density, likely corresponding to the Spt16 MD. This suggests that FACT is not fully engaged in the nucleosome at this stage, although it is tethered to the exposed H2A-H2B through the Spt16 CTD.

EC49 reveals downstream nucleosome unfolding and FACT engagement

The EC49 structure reveals the state in which the EC has further advanced, and its leading edge is located just before SHL(0) (dyad) within the downstream nucleosome (Fig. 3A, middle, and fig. S17). An ~70-bp DNA segment [SHL(−7) ~ SHL (−1)] is peeled from the histone octamer surface, and most of the DNA interactions with the proximal H3-H4 are lost (Fig. 4C and fig. S16). An ~30-bp DNA segment [SHL(+7) ~ SHL(+5)] is also removed from the distal end of the nucleosome, and the DNA-binding surface of the distal H2A-H2B is partly exposed. Thus, the histone octamer in EC49 is wrapped by only an ~55-bp DNA segment. The distal DNA dissociation is apparently due to steric hindrance between the DNA and the EC, because modeling of the DNA causes clashes with Elf1 and the RNAPII clamp head. The trajectory of the distal DNA would also clash with the proximal DNA because it becomes straight due to its removal from the proximal H3-H4 (Fig. 4I), suggesting that the removal of the proximal half of the DNA itself could destabilize the distal DNA binding to histones.

In EC49, clear density was observed for the main body of FACT, which associates with the considerably unwrapped nucleosome (Fig. 4C). This is consistent with the previous report that FACT associates with RNAP-destabilized nucleosomes (*37*). The Spt16 MD of FACT interacts with the H3-H4 tetramer and the proximal H2A-H2B dimer and becomes integrated with the histone hexamer unit. The densities for the H2A-docking domain (residues 84 to 119) and the H3 N-terminal helix (residues 45 to 55) are missing, and they are displaced by the middle part of the Spt16 MD (compare Fig. 4D with Fig. 4, B and G, and fig. S18B). The Spt16 residues V831, T894, and L873 form hydrophobic contacts with I112, L109, and V101 of H3, respectively. The HA1 helix and its flanking loop of the Spt16 MD (residues 780 to 812) are highly acidic and directly bind the exposed basic surface of the proximal H3-H4, around

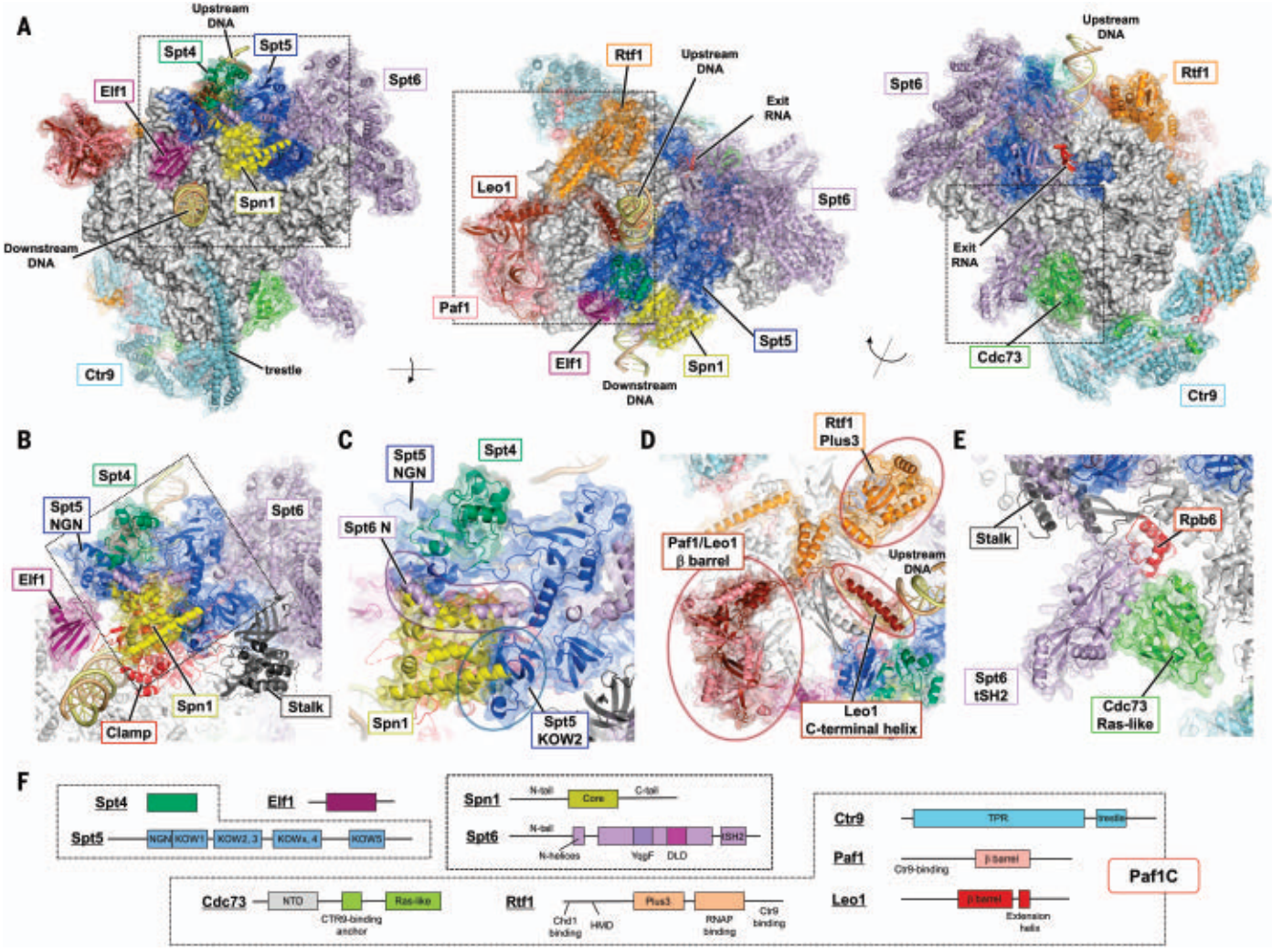


Fig. 2. Cryo-EM structures of the RNAPII EC. (**A**) Overall structures of the EC shown from downstream (left), upstream (middle), and the RNA-exit site (right). RNAPII is shown as a white surface model. The RNAPII-associated elongation factors are shown as ribbon models overlaid with transparent surface models. (**B**) Close-up view of the downstream elongation factors corresponding to the dotted rectangle in (A) (left). The RNAPII clamp and stalk are colored red and black, respectively. (**C**) Close-up view on the interactions around Spn1. The view corresponds to the dotted rectangle in (B). (**D**) Close-up view around the Paf1/Leo1 β barrel corresponding to the dotted rectangle in (A) (middle). (**E**) The Cdc73 Ras-like domain and the Spt6 tSH2 domain. The view corresponds to the dotted rectangle in (A) (right). The RNAPII stalk and Rpb6 are colored black and red, respectively. (**F**) Domain architectures of elongation factor subunits.

the H3 α1 helix (residues 63 to 69) and the H4 N-terminal tail region (residues 24 to 32) (fig. S18C). Moreover, Spt16 has an additional α helix (linker helix, residues 484 to 516), which was not modeled in the previous FACT-nucleosome structures (*36*) (Fig. 4C). This helix is located between the NTD and DD in the primary sequence of Spt16 (Fig. 3C) but is integrated within the MD in the tertiary structure (Fig. 4D). This helix extends along the α2 helix of H2A, and the Spt16 R484 residue at the tip of the helix could interact with nucleosomal acidic patch residues (E61 in H2A and E105 in H2B), in good agreement with previous reports (*38*, *39*). The Spt16 CTD tail also interacts with the exposed surface of the proximal H2A-H2B. Thus, Spt16 disrupts the important interaction between H2A-H2B and H3-H4 mediated by the H2A-docking domain (*40*) while stabilizing the subnucleosomal structure through its linker helix, MD, and CTD. This is consistent with reports that FACT prevents the loss of the proximal H2A-H2B dimer (*16*, *41*).

The Spt16-histone interaction observed in EC49 is different from that in the previously reported structure of the nucleosome-transcribing RNAPII bound with FACT, in which the leading edge of RNAPII is located at SHL(−4) (*35*) (Fig. 4H, left). In this SHL(−4) structure, the Spt16 MD is not engaged with the histones, because the proximal H3-H4 is still covered by DNA, which would hinder the access of the Spt16 HA1 helix to H3-H4 (Fig. 4H, right, and fig. S16). Instead, the binding mode of the Spt16 MD to the H3-H4 tetramer in EC49 is rather similar to that observed in the crystal structure of the human Spt16 MD complexed with the DNA-free H3-H4 tetramer (*42*) (Fig. 4H, middle, right). This suggests that the FACT Spt16 is only fully engaged with the nucleosome when the EC reaches SHL(0) to remove nearly half of the nucleosomal DNA from the histone octamer, before the detachment of the FACT-histone intermediate complex from the DNA and its subsequent transfer.

The other FACT subunit, Pob3, is located on the opposite side of the nucleosome, and its MD contacts the nucleosomal DNA around SHL(+1) (Fig. 4C). The Pob3 DD and Spt16 DD form a domain dimer (FACT DD), which sits on the nucleosomal DNA around SHL(+0.5). These domain locations are similar to those in the SHL(−4) structure (*35*), but their orientations

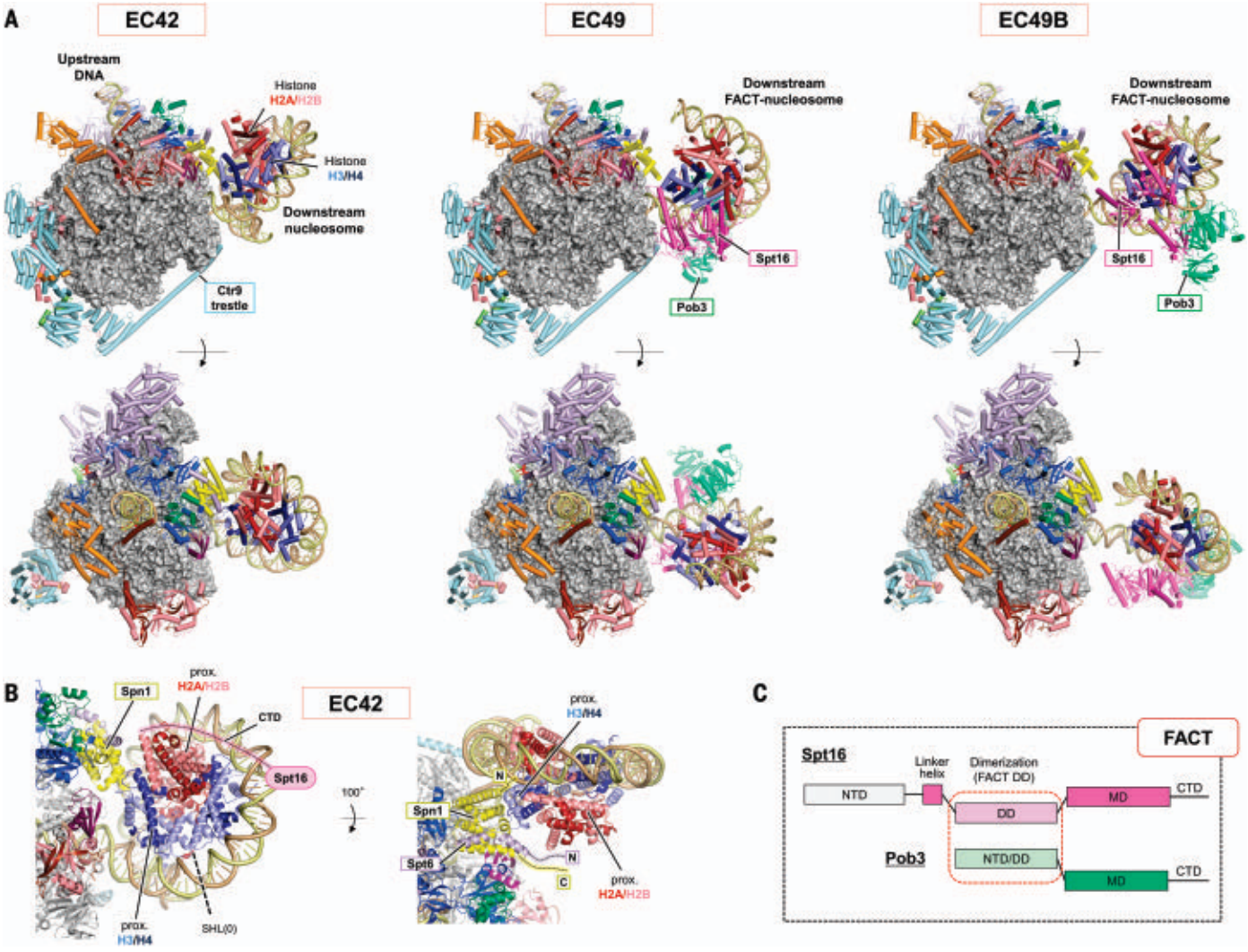


Fig. 3. Cryo-EM structures of EC-downstream nucleosome complexes. (**A**) Overall structures of the EC42, EC49, and EC49B complexes. RNAPII is shown as a white surface model. Other proteins and nucleic acids are shown in cartoon representation. (**B**) Close-up view of the EC42 downstream nucleosome and the associating factors shown in cartoon representation. (**C**) Domain architecture of FACT.

are different, probably because of the distinct histone-binding modes of the Spt16 MD (fig. S16).

EC49B reveals downstream shifting of the FACT-nucleosome complex

In EC49B, the EC is stalled at the same position on the template DNA as in EC49. However, the nucleosome is shifted downstream by ~17 bp, and the EC and FACT-nucleosome complex are far apart (Fig. 3A, right, and figs. S1B and S17). In the nucleosome, the histone octamer is rewrapped with DNA to SHL(−1.5) (Fig. 4E). The Spt16 MD is detached from the histones, because the DNA rewrapping around SHL(−1) is incompatible with Spt16 MD binding to the proximal H3-H4. Because of this change, the H2A-docking domain and the H3 N-terminal helix became visible (Fig. 4F). By contrast, the density map reveals that the Spt16 CTD tail remains bound to the proximal H2A-H2B (Fig. 4E and fig. S18D). The FACT DD is repositioned and sits on the nucleosomal dyad, and the Pob3 MD is also repositioned and binds the nucleosomal DNA at SHL (0 to +0.5). A weak density is observed for the putative Pob3 CTD tail, which interacts with the DNA-peeled surface of the distal H2A-H2B and the αN helix of the distal H3 (fig. S18D). Overall, these FACT-nucleosome interactions are similar to those in the previously reported FACT-nucleosome complex structure (*36*) (fig. S16).

The Spt16 engagement with the histones and the removal of most of the nucleosomal DNA observed in EC49 are likely to facilitate the release of the FACT-histone intermediate, which can be transferred to other DNA regions either downstream or upstream of the EC. The EC49B structure probably originated from the downstream "jump" of the FACT-histone intermediate.

Upstream nucleosome reassembly

EC58^{hex} reveals deposition of the histone hexamer held by FACT on the upstream side of EC

In EC58^{hex}, the EC leading edge overruns the dyad position of the original downstream nucleosome. EC58^{hex} no longer has a nucleosome density downstream of the EC, but instead shows a density for a subnucleosome structure upstream of the EC (Fig. 5A, fig. S19, and movies S2 and S3). A histone hexamer held by FACT is deposited onto DNA at the DNA exit site of the EC. This position is ~45 bp upstream of the original nucleosome position (fig. S1B). The resultant hexasome is only wrapped by an ~40-bp DNA segment, covering SHL(−3) to SHL(0), representing an early stage of the upstream nucleosome reassembly (Fig. 5B). Spt4, Spt5 (KOW1 domain), Spt6 (YqgF and DLD domains), Leo1 (C-terminal helix), and Rtf1 (Plus3 domain) form the rim of the DNA

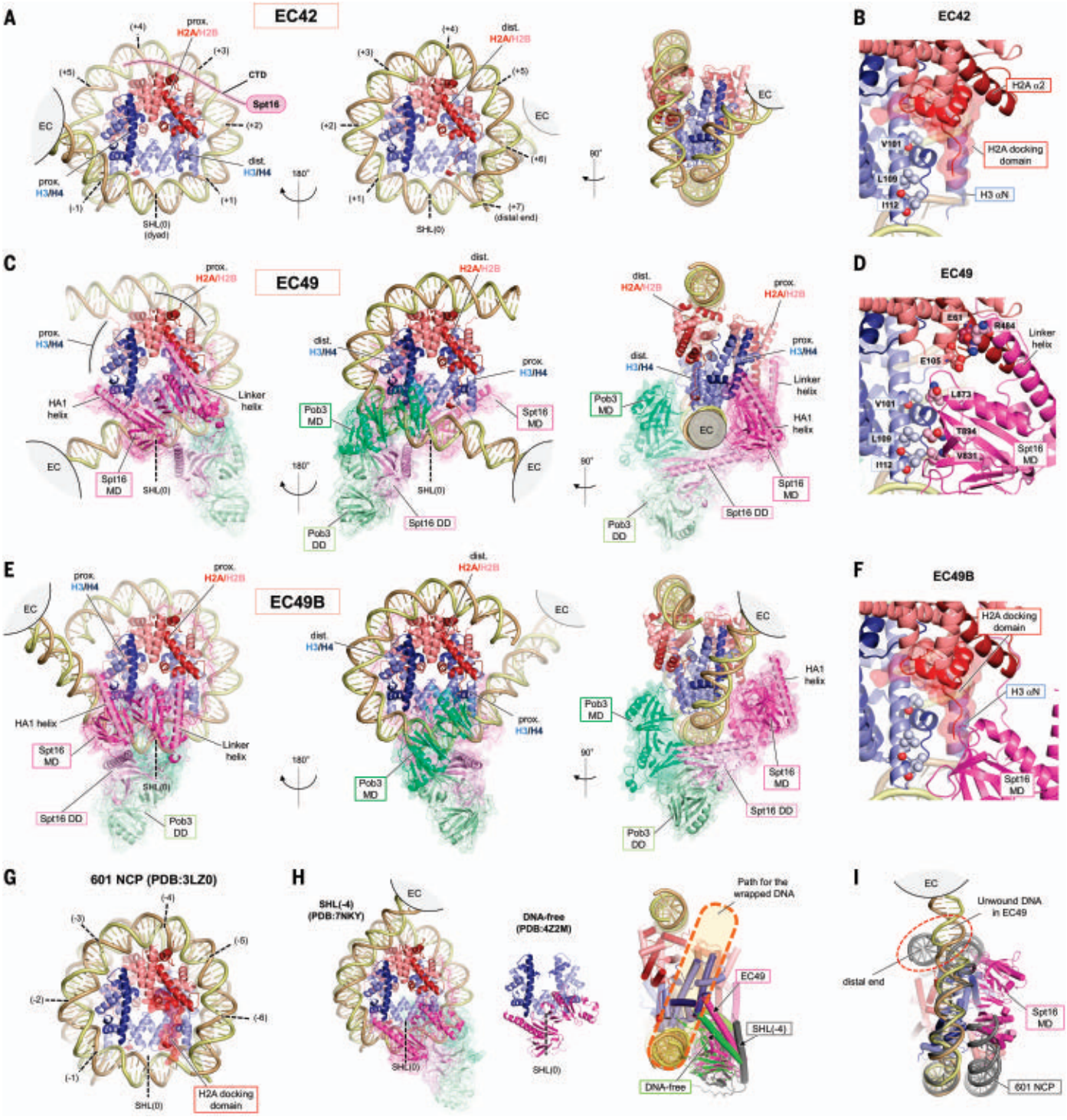


Fig. 4. Structures of the downstream nucleosome. (**A**) Structure of the EC42 nucleosome shown from the Spt16 side (left), the Pob3 side (middle), and the EC side (right). SHLs are indicated with dashed lines. (**B**) Interface between the proximal H2A-H2B and the distal H3-H4 in EC42. A red transparent surface is overlaid on the H2A-docking domain. (**C**) Structure of the EC49 FACT-nucleosome complex. Transparent surfaces are overlaid on the Spt16 and Pob3 cartoon models. (**D**) Binding interface of Spt16 MD to histones observed in EC49 with the missing H2A-docking domain. The view is same as in (B). The interface residues are shown as spheres. (**E**) Structure of the EC49B FACT-nucleosome complex. (**F**) The H2A-docking domain and the detached Spt16 MD observed in EC49B. (**G**) Crystal structure of the intact 601 nucleosome [PDB: 3LZ0 (*61*)]. A red transparent surface is overlaid on the H2A-docking domain. SHLs 0 to −6 are indicated with dashed lines. (**H**) Structural comparisons of the SHL(−4)-FACT-nucleosome complex [PDB: 7NKY (*35*), left], the DNA-free Spt16-$(H3\text{-}H4)_2$ complex [PDB: 4Z2M (*42*), middle], and the EC49 nucleosome in this study. In the right panel, the three structures are overlaid in the same orientation as in (C) (right), and their Spt16 MD HA1 helices are indicated by arrows. The location of the wrapped DNA in the SHL(−4) structure is indicated by an orange dashed line, which interferes with the Spt16 MD binding. (**I**) Superimposition of the EC49 nucleosome with the intact nucleosome showing the incompatibility of the nucleosome core particle (NCP) distal end and the unwound DNA in EC49.

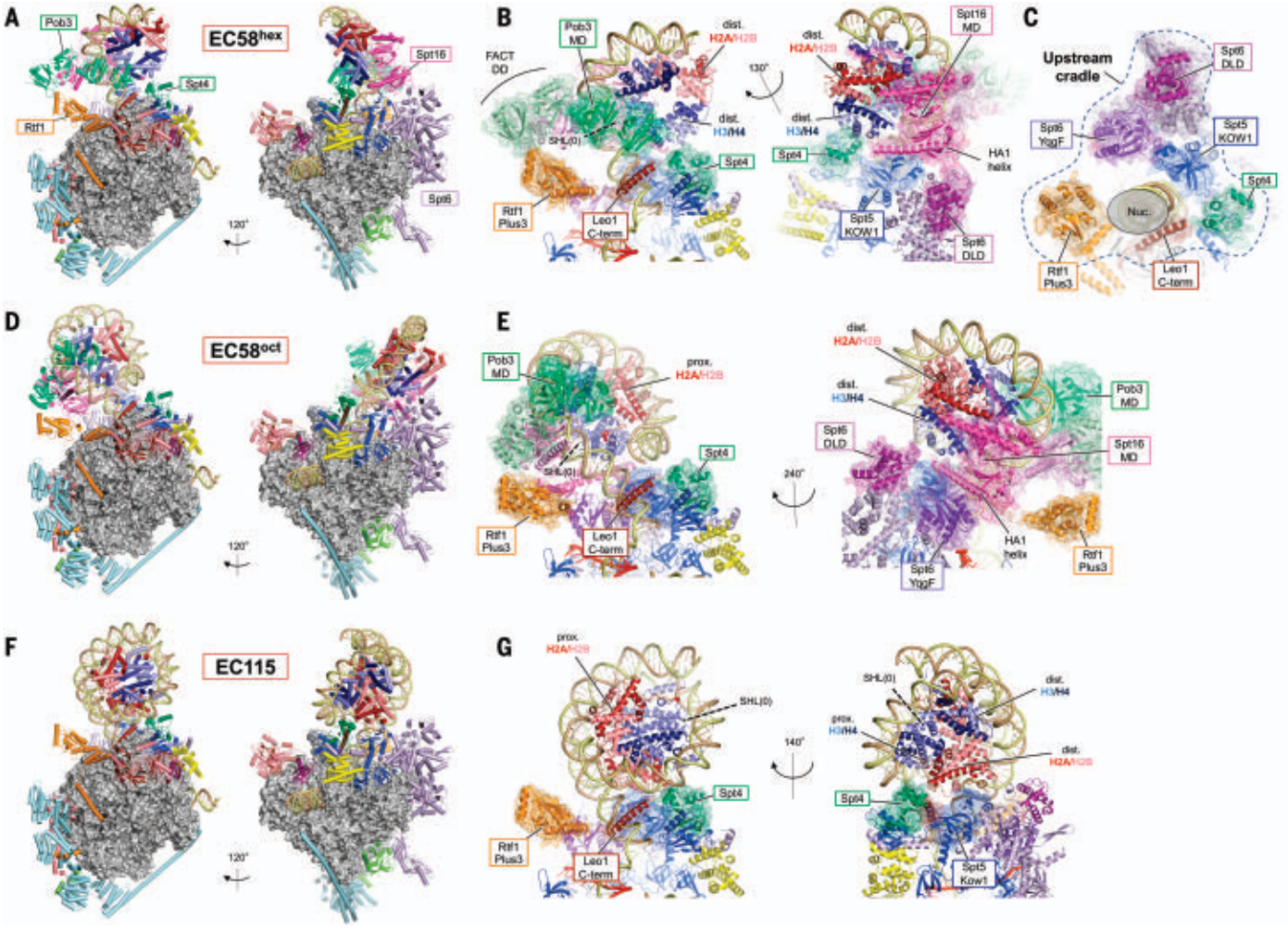


Fig. 5. Cryo-EM structures of EC-upstream nucleosome complexes. (**A**) Overall structure of the EC58^{hex} complex. (**B**) Interactions between the FACT-subnucleosome and the elongation factors in EC58^{hex}. The elongation factor domains constituting the cradle and the FACT subunits are shown as ribbon models with transparent surfaces, with RNAPII removed for clarity. (**C**) Close-up view of the upstream cradle shown from the nucleosome side. (**D**) Overall structure of the EC58^{oct} complex. (**E**) Interactions between the FACT-subnucleosome and the elongation factors in EC58^{oct}, with RNAPII removed for clarity. (**F**) Overall structure of the EC115 complex. (**G**) Interactions between the nucleosome and the elongation factors in EC115, with RNAPII removed for clarity.

exit and serve as a "cradle" to accommodate the FACT-hexasome complex (Fig. 5, B and C). The Spt16 MD contacts the Spt5 KOW1 domain and the Spt6 DLD domain, and the FACT DD contacts the Rtf1 Plus3 domain (Fig. 5B and fig. S20B). The exposed face of the distal H3 directly contacts the acidic face of Spt4.

In the FACT-hexasome complex, the histone hexamer is sandwiched between the Spt16 and Pob3 MDs in a unique binding mode (Fig. 5B and fig. S16). The binding manner of the Spt16 MD to the histone hexamer is similar to that in EC49. The HA1 helix of the Spt16 MD interacts with the exposed basic surface of the distal H3-H4, confirming that this engagement depends on the lack of H3-H4 interactions with DNA around SHL(−1). By contrast, the Pob3 MD interacts differently with the histone hexamer compared with that in EC49. It directly contacts the H3-H4 tetramer surface, which partially overlaps with the H2A-H2B–binding site (Fig. 6). The Pob3 MD binding to H3-H4 probably occurs because the histone hexamer lacks the proximal H2A-H2B, and thus there is no steric hindrance between the Pob3 MD and H2A-H2B. This unique Pob3-hexasome binding mode is similar to that in the previously reported human FACT-hexasome complex structure (*36*) (fig. S16). The FACT DD in EC58^{hex} resides in an intermediate position between those in EC49 and the human complex structure, probably because of steric restrictions caused by the Spt16 MD and Pob3 MD positions. We presume that the observed structure of the histone hexamer held between the Spt16 and Pob3 MDs represents the structure of the putative FACT-histone transfer intermediate.

EC58^{oct} with the FACT-octasome complex on the EC upstream side

In EC58^{oct}, the EC is located at the same DNA position as that in EC58^{hex} but reveals a subnucleosome structure (octasome) containing a histone octamer held by FACT on the EC upstream side (Fig. 5D and fig. S19). EC58^{oct} likely represents the state in which the proximal H2A-H2B dimer is further deposited onto the histone hexamer in EC58^{hex}. An ~75-bp DNA segment is wrapped around the histone octamer, encompassing SHL(−6) to SHL(+0.5) (Fig. 5E and fig. S16). The FACT-octasome in EC58^{oct} is similar to that in EC49, although the entire surface of the proximal H2A-H2B is covered by DNA in EC58^{oct}. In contrast to EC58^{hex}, the Pob3 MD is detached from the H3-H4 tetramer surface in EC58^{oct}, probably due to steric hindrance with the added proximal H2A-H2B (Fig. 6). This suggests that H3-H4 binding by the Pob3 MD and by the H2A-H2B are mutually exclusive.

The FACT-subnucleosome orientations relative to the EC are different between EC58^{hex} and EC58^{oct}. In EC58^{oct}, the upstream DNA is

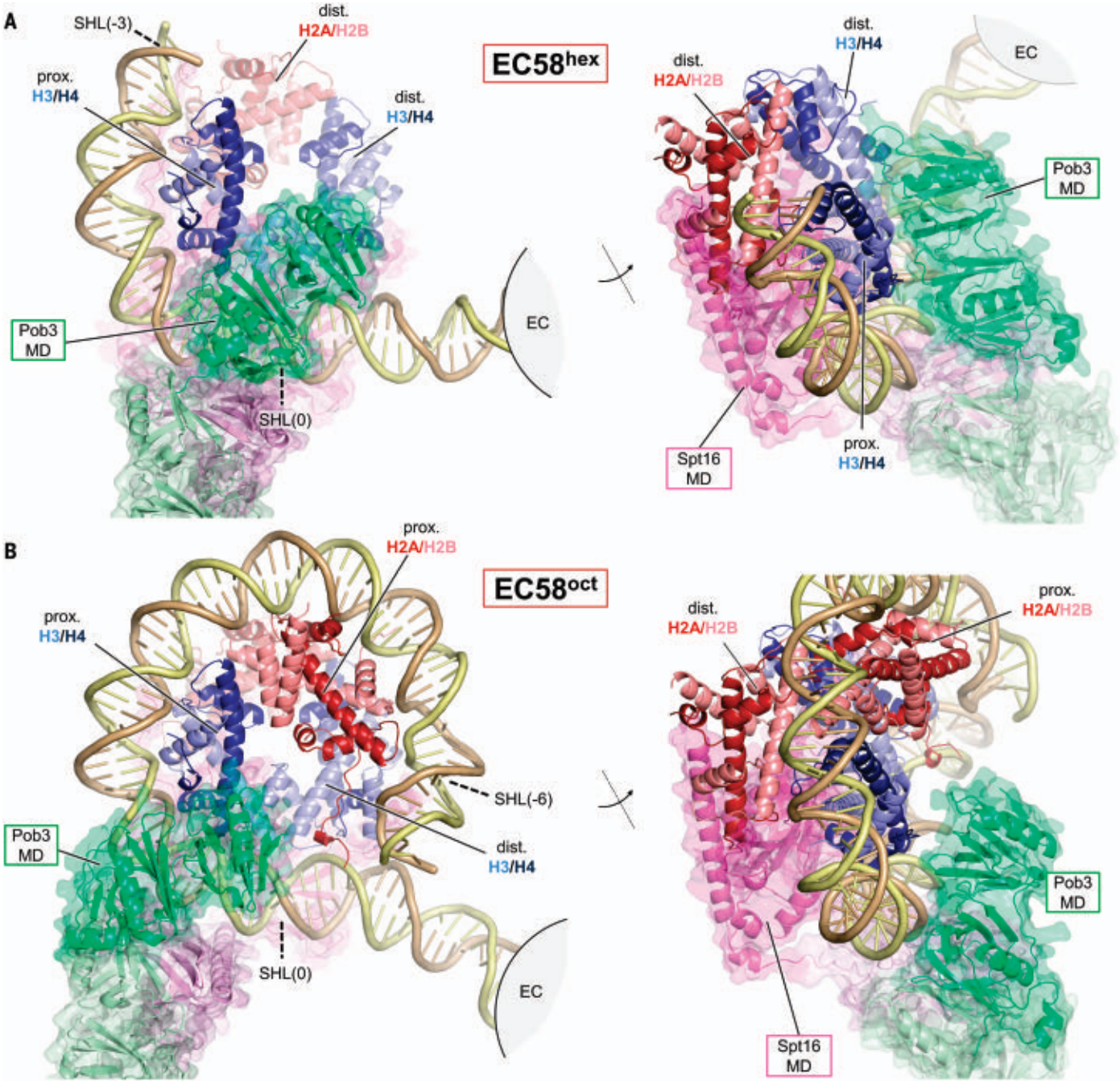


Fig. 6. Structures of the upstream subnucleosomes. (**A**) Structure of the hexasome associated with FACT in EC58^{hex} in two different orientations. (**B**) Structure of the octasome associated with FACT in EC58^{oct} in two different orientations.

slightly bent and the FACT-octasome complex is rotated by ~70° relative to the FACT-hexasome complex in EC58^{oct} (Fig. 5E and fig. S20A). This change is due to the hexasome to octasome conversion. Because the nucleosomal DNA covers the proximal H2A-H2B dimer in EC58^{oct}, the octasome would need to change its orientation to avoid steric clashes of the DNA with Spt4 and the Leo1 C-terminal helix. This switches the interaction manner between the FACT-subnucleosome complex and the EC cradle (fig. S20B). The Spt16 MD binds the Spt6 YqgF and Spt5 KOW1 domains, mainly through its HA1 helix. The exposed distal H3-H4 is close to the Spt6 DLD domain. The FACT DD maintains the contact with Rtf1 Plus3, which changes its location according to the FACT reorientation.

EC115 with a near-complete nucleosome on the upstream side of EC

The leading edge of EC115 resides in a DNA position near the exit of the original downstream nucleosome. EC115 reveals a nearly complete nucleosome structure on the upstream side of the EC (Fig. 5F and fig. S19). The nucleosome resides ~15 bp downstream of that in EC58^{hex}/EC58^{oct}, suggesting that the nucleosome has adapted to a more favorable DNA region (fig. S1B). The histone octamer is covered by an ~120-bp DNA segment from SHL(−7) to SHL(+4.5) (fig. S16). No density for FACT is visible, indicating that FACT has already dissociated from the nucleosome at this stage. This is probably because the distal H3-H4 and nearly half of

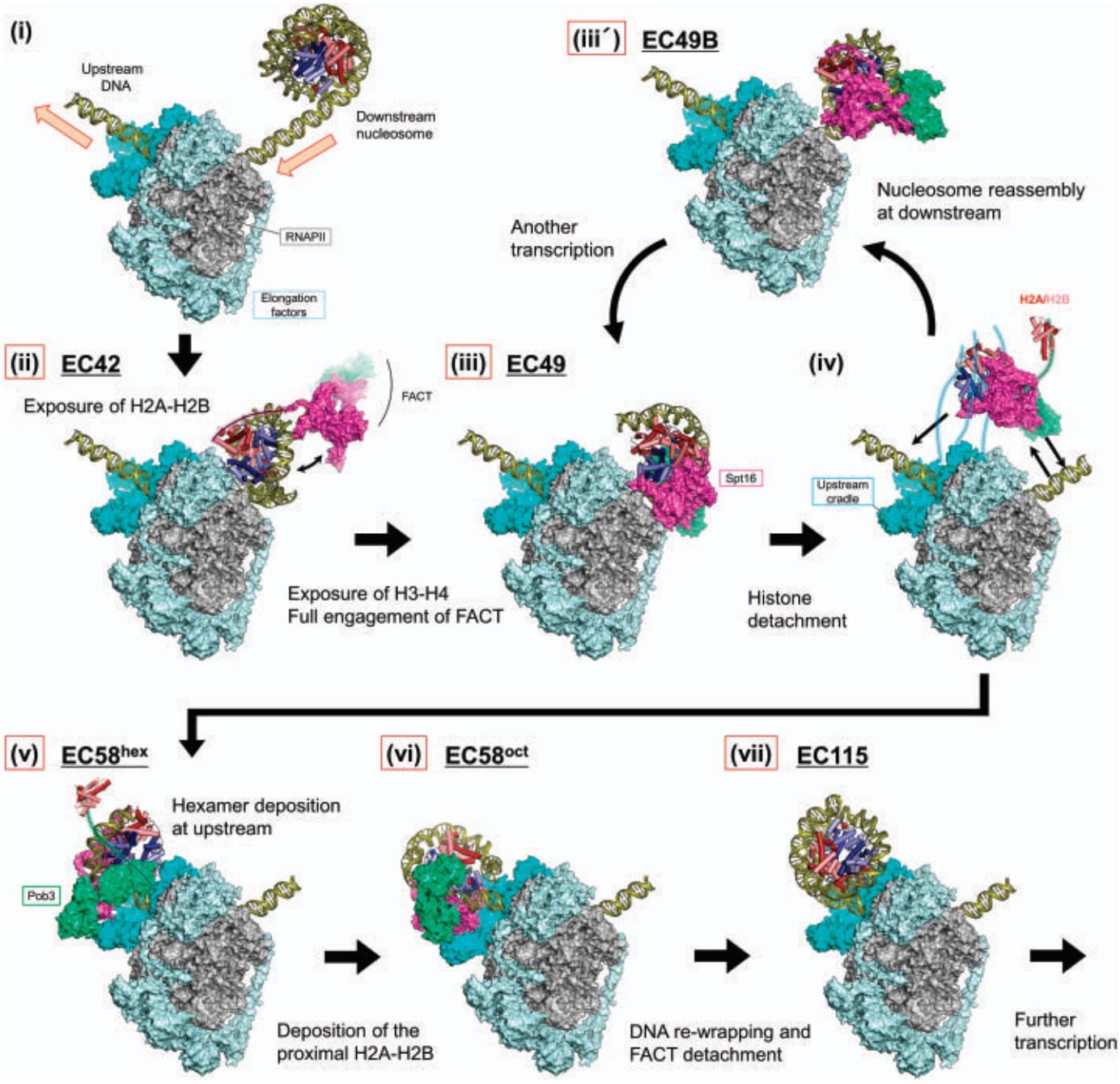


Fig. 7. Schematic model of transcription over a nucleosome mediated by EC and FACT. Models of (ii), (iii), (iii´), (v), (vi), and (vii), as marked with red rectangles, are based on the cryo-EM structures of EC42, EC49, EC49B, EC58^{hex}, EC58^{oct}, and EC115, respectively. The FACT-histone complex in (iv) was modeled based on that in EC58^{hex} without DNA.

the distal H2A-H2B are rewrapped by DNA, and its presence is incompatible with the Spt16 MD and CTD binding. The remaining exposed surface of the distal H2A-H2B contacts the acidic face of Spt4 and the Spt5 KOW1 domain in the cradle (Fig. 5G). As mentioned above, Spt4 also interacts with the exposed H3-H4 in EC58^{hex}, suggesting its histone chaperone-like role. This might be relevant to the observation that Spt4 influences the positions and spacing of gene-body nucleosomes (*43*). The Leo1 C-terminal helix is also close to H2A-H2B. Thus, the cradle flexibly adapts to the dynamically changing structures and orientations of the subnucleosome intermediates formed in EC58^{hex}, EC58^{oct}, and EC115, and may provide a platform to support the sequential nucleosome reassembly process upstream of the EC (fig. S20B). Finally, we performed transcription on temp115 in the absence of Spt6, Spn1, Paf1C, and FACT and analyzed the formed complex by cryo-EM. No clear nucleosome structure was observed either upstream or downstream of RNAPII, indicating the significance of these factors in the upstream nucleosome reassembly (fig. S21).

Discussion

In the present study, we performed nucleosome transcription by RNAPII in the presence of transcription elongation factors and histone chaperones and obtained snapshots

of nucleosome disassembly followed by reassembly according to the RNAPII progression. RNAPII forms an EC with Spn1, Spt6, Elf1, Spt4/5, and Paf1C, which compose the nucleosome disassembly face downstream of the EC and the reassembly face (cradle) upstream of the EC.

On the basis of the obtained structural snapshots, we propose the following mechanism by which the EC cooperates with FACT to mediate the histone transfer for the EC traversal through a nucleosome (Fig. 7). First, the EC approaches the downstream nucleosome to start unwrapping DNA from the proximal histones (i). Second, once the proximal H2A-H2B is exposed, FACT begins to access the histones through the Spt16 CTD tail (ii). At this stage, the main body of FACT does not stably bind the nucleosome. Third, as the EC edge approaches the nucleosomal dyad [SHL(0)], removing DNA from the proximal H3-H4, Spt16 is fully engaged with the histones (iii). The Spt16 MD and the linker helix clutch the histone hexamer unit consisting of the H3-H4 tetramer and the proximal H2A-H2B dimer. At this stage, the distal DNA is simultaneously unwrapped, and only a third of the histone octamer surface is covered by DNA. Fourth, further EC advancement causes histone detachment from the DNA for the subsequent histone transfer to the upstream DNA (iv). FACT holds the histone hexamer with its main body and presumably tethers the dissociated distal H2A-H2B dimer through the Pob3 CTD (*44*), although the latter might be less efficient, as indicated by the loss of a single H2A-H2B dimer in vitro (*45*) and the faster H2A-H2B exchange over H3-H4 in vivo (*46*, *47*). We speculate that the tail domains of Spn1, Spt6, Elf1, Spt5, Rtf1, Leo1, and Ctr9 (*32*, *48*), which are not observed in the current structures, could interact with the detached FACT-histone intermediate to prevent its diffusion, for its efficient upstream transfer. Alternatively, the FACT-histone octamer complex can be transferred downstream (iii′). When the EC catches up with the downstream-shifted nucleosome, it would help to regenerate the FACT-histone intermediate for its upstream transfer. Fifth, the cradle on the upstream side of the EC captures the histone hexamer held by FACT, which deposits the histone hexamer on the DNA exiting from the EC to form the hexasome (v). Sixth, the remaining H2A-H2B is deposited on the hexasome to reassemble the octasome (vi). Finally, further EC progression extends the exiting upstream DNA, which rewraps the distal H3-H4, causing FACT dissociation (vii). The EC cradle adapts to the series of upstream intermediates, supporting the nucleosome reassembly process. Further transcription would lead to rewrapping of the distal H2A-H2B and the formation of a complete nucleosome behind the EC. This model explains the mechanism by which the nucleosomal histones and histone modifications are preserved in the wake of the transcription machinery.

The (sub)nucleosome structures in EC58^{hex}, EC58^{oct}, and EC115 are shifted upstream relative to the initial nucleosome position, suggesting that the transcription-coupled histone transfer from the EC downstream to the upstream is accompanied by a nucleosome upstream shift. The magnitude of the shift would vary, depending on the DNA sequence context, the EC elongation rate and the pause probability. In fact, the nucleosome position in EC115 is shifted ~15 bp downstream relative to that in EC58^{hex}/EC58^{oct}. In addition, if the upstream histone transfer occurs after the nucleosome is shifted downstream, as observed in EC49B, then the net upstream shift becomes smaller. Chromatin remodelers, such as Chd1 and Isw1, might move the reassembled nucleosome back to the original position to restore the nucleosome positions and spacing in chromatin (*2*, *49*). It was previously proposed that RNAPII alone can pass through a nucleosome through a looped-template intermediate ("Ø loop") without shifting the nucleosome position (*45*, *50*, *51*). A recently reported structure of an RNAPII-nucleosome complex contained an approximately 90-bp looped-template DNA, which was proposed to mediate the nucleosome retention (*52*). These Ø-loop and 90-bp-loop models are likely to differ from our current mechanism, which relies on elongation factors and histone chaperones.

The structures of EC49, EC58^{hex}, and EC58^{oct} reveal that the Spt16 MD becomes a common interface between the EC and subnucleosome structures either downstream or upstream of the EC. The HA1 helix of the Spt16 MD always faces the EC, and thus the proximal histones in the downstream nucleosome could become the distal histones in the upstream nucleosome when they are transferred from downstream to upstream of the EC. This suggests that the traverse across the nucleosome by the EC is accompanied by the flipping of the histone octamer, implying that transcription reverses the histone orientations in gene-body nucleosomes.

The current structures of the EC and the downstream and upstream (sub)nucleosomes provide a solid foundation for future studies on chromatin transcription, transcription-coupled histone modification, chromatin remodeling, DNA repair, and more. Nucleosome disassembly and reassembly also accompany DNA replication, in which FACT could mediate histone transfer in a similar way to transcription. These aspects are subjects for future studies.

Methods Summary

We previously established a method to analyze the nucleosome-transcribing RNAPII structures by combining in vitro transcription of a nucleosomal DNA template by a purified enzyme and cryo-EM analysis of the RNAPII-nucleosome complex species formed during the reaction (*32*, *53*). In the present study, to capture structural snapshots of EC traversing the nucleosome, three nucleosomal DNA templates were designed (Fig. 1, A and B): temp42/58, containing two thymidine (T) tracts at positions 42 and 58 relative to the nucleosome entry and a T-less region from the transcription start position to the first T tract (*53*), and temp49 and temp115, containing a T tract at positions 49 and 115, respectively. Assuming that the RNAPII catalytic sites are at positions 42, 49, 58, and 115, the RNAPII leading edge is expected to reach SHL(−1), SHL(0) (dyad), SHL(+1), and SHL(+6), respectively.

Transcription was conducted in the yeast *Komagataella pastoris* RNAPII with the transcription elongation factors Spt6, Spn1, Spt4/5, Elf1, Paf1C, and TFIIS and the histone chaperone FACT. The reaction also contained P-TEFb and casein kinase II (CK2). P-TEFb phosphorylates Spt4/5, Paf1C, and the CTD of the RNAPII Rpb1 subunit, promoting Spt6 and Paf1C binding to RNAPII (*26*, *54*). CK2 phosphorylates FACT, Spt6, Elf1, Paf1C, and histone H2A and facilitates FACT-nucleosome and Spt6-Spn1 interactions (*24*, *55*–*59*). The transcription experiment showed that Spt6, Spn1, Spt4/5, Elf1, and Paf1C enhance the RNAPII processivity through the nucleosomal template (fig. S1A). FACT addition generated the RNAPII fractions stalled beyond the nucleosomal dyad, as observed previously (*16*). For the cryo-EM sample preparation, transcription was conducted with all of these factors on temp42/58, temp49, and temp115 in the presence of 3′-deoxyadenosine triphosphate (Fig. 1A). Transcription was stalled at the expected T positions within the three templates (Fig. 1C). For temp42/58, RNAPII stalled at the 42nd and 58th T positions, the latter of which was due to the readthrough of the first T tract by a certain fraction of RNAPII. The complex species formed in each transcription experiment were cross-linked and partially purified by the gradient fixation (GraFix) method (*60*) for structure determination by cryo-EM single-particle analysis.

For cryo-EM analyses, 54741, 35110, and 33987 micrographs were collected from the samples prepared with temp42/58, temp49, and temp115, respectively (table S1). Initial data processing revealed that these samples are highly heterogeneous in terms of the nucleosome arrangement relative to RNAPII and the presence or absence of protein factors. Therefore, we first focused on reconstructing the RNAPII EC map using the temp42/58 dataset while masking out the nucleosomes (fig. S3). In the consensus reconstruction, the densities for the elongation factors, especially

those on the periphery of the complex, are of lower quality because of their flexibility. These densities were improved by focused classifications. By combining seven locally refined maps, we built an atomic model of the RNAPII EC containing Spn1, Spt6, Spt4/5, Elf1, and Paf1C (figs. S4 to S6).

Starting from the consensus EC particles, extensive three-dimensional classifications were performed to determine the EC structures bound with a nucleosome (Fig. 1D, figs. S7 to S13, and table S2). The temp42/58 dataset yielded three complexes: a complex in which the EC stalls at the 42nd T position (EC42) and two complexes in which the EC stalls at the 58th T position ($EC58^{hex}$ and $EC58^{oct}$). The temp49 and temp115 datasets yielded two complexes at the 49th T position (EC49 and EC49B) and a complex at the 115th position (EC115). EC42 contains a nucleosome that is partially associated with FACT at the downstream edge of the EC. EC49 and EC49B reveal a substantially unwrapped nucleosome associated with FACT downstream of the EC. By contrast, $EC58^{hex}$ and $EC58^{oct}$ represent subnucleosome structures at the upstream edge of the EC and contain a histone hexamer and octamer bound with FACT, respectively. EC115 reveals a near-complete nucleosome located upstream of the EC, but no density for FACT is visible in this structure. More details are described in the supplementary materials.

REFERENCES AND NOTES

1. H. Kwak, J. T. Lis, Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508 (2013). doi: 10.1146/annurev-genet-110711-155440; pmid: 24050178
2. S. Venkatesh, J. L. Workman, Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015). doi: 10.1038/nrm3941; pmid: 25650798
3. T. Kujirai, H. Kurumizaka, Transcription through the nucleosome. *Curr. Opin. Struct. Biol.* **61**, 42–49 (2020). doi: 10.1016/j.sbi.2019.10.007; pmid: 31790919
4. M. J. Rossi *et al*., A high-resolution protein architecture of the budding yeast genome. *Nature* **592**, 309–314 (2021). doi: 10.1038/s41586-021-03314-8; pmid: 33692541
5. P. Cramer, Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019). doi: 10.1038/s41586-019-1517-4; pmid: 31462772
6. L. Muniz, E. Nicolas, D. Trouche, RNA polymerase II speed: A key player in controlling and adapting transcriptome composition. *EMBO J.* **40**, e105740 (2021). doi: 10.15252/embj.2020105740; pmid: 34254686
7. N. I. Reim *et al*., The conserved elongation factor Spn1 is required for normal transcription, histone modifications, and splicing in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **48**, 10241–10258 (2020). doi: 10.1093/nar/gkaa745; pmid: 32941642
8. A. M. Francette, S. A. Tripplehorn, K. M. Arndt, The Paf1 complex: A keystone of nuclear regulation operating at the interface of transcription and chromatin. *J. Mol. Biol.* **433**, 166979 (2021). doi: 10.1016/j.jmb.2021.166979; pmid: 33811920
9. T. M. Decker, Mechanisms of transcription elongation factor DSIF (Spt4-Spt5). *J. Mol. Biol.* **433**, 166657 (2021). doi: 10.1016/j.jmb.2020.09.016; pmid: 32987031
10. G. Orphanides, G. LeRoy, C.-H. Chang, D. S. Luse, D. Reinberg, FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105–116 (1998). doi: 10.1016/S0092-8674(00)80903-4; pmid: 9489704
11. T. Formosa, F. Winston, The role of FACT in managing chromatin: Disruption, assembly, or repair? *Nucleic Acids Res.* **48**, 11929–11941 (2020). doi: 10.1093/nar/gkaa912; pmid: 33104782
12. R. Belotserkovskaya *et al*., FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**, 1090–1093 (2003). doi: 10.1126/science.1085703; pmid: 12934006
13. P. B. Mason, K. Struhl, The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol. Cell. Biol.* **23**, 8323–8333 (2003). doi: 10.1128/MCB.23.22.8323-8333.2003; pmid: 14585989
14. A. Jamai, A. Puglisi, M. Strubin, Histone chaperone spt16 promotes redeposition of the original h3-h4 histones evicted by elongating RNA polymerase. *Mol. Cell* **35**, 377–383 (2009). doi: 10.1016/j.molcel.2009.07.001; pmid: 19683500
15. C. Jeronimo, C. Poitras, F. Robert, Histone recycling by FACT and Spt6 during transcription prevents the scrambling of histone modifications. *Cell Rep.* **28**, 1206–1218.e8 (2019). doi: 10.1016/j.celrep.2019.06.097; pmid: 31365865
16. F. K. Hsieh *et al*., Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7654–7659 (2013). doi: 10.1073/pnas.1222198110; pmid: 23610384
17. M. B. Ardehali *et al*., Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J.* **28**, 1067–1077 (2009). doi: 10.1038/emboj.2009.56; pmid: 19279664
18. A. Narain *et al*., Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol. Cell* **81**, 3110–3127.e14 (2021). doi: 10.1016/j.molcel.2021.06.016; pmid: 34233157
19. K. Žumer *et al*., Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Mol. Cell* **81**, 3096–3109.e8 (2021). doi: 10.1016/j.molcel.2021.05.028; pmid: 34146481
20. A. Bortvin, F. Winston, Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* **272**, 1473–1476 (1996). doi: 10.1126/science.272.5267.1473; pmid: 8633238
21. C. D. Kaplan, L. Laprade, F. Winston, Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**, 1096–1099 (2003). doi: 10.1126/science.1087374; pmid: 12934008
22. S. M. Doris *et al*., Spt6 is required for the fidelity of promoter selection. *Mol. Cell* **72**, 687–699.e6 (2018). doi: 10.1016/j.molcel.2018.09.005; pmid: 30318445
23. N. J. Krogan *et al*., RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: A targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992 (2002). doi: 10.1128/MCB.22.20.6979-6992.2002; pmid: 12242279
24. D. Prather, N. J. Krogan, A. Emili, J. F. Greenblatt, F. Winston, Identification and characterization of Elf1, a conserved transcription elongation factor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 10122–10135 (2005). doi: 10.1128/MCB.25.22.10122-10135.2005; pmid: 16260625
25. J. A. Fischbeck, S. M. Kraemer, L. A. Stargell, SPN1, a conserved gene identified by suppression of a postrecruitment-defective yeast TATA-binding protein mutant. *Genetics* **162**, 1605–1616 (2002). doi: 10.1093/genetics/162.4.1605; pmid: 12524336
26. S. M. Vos *et al*., Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018). doi: 10.1038/s41586-018-0440-4; pmid: 30135578
27. S. M. Vos, L. Farnung, A. Linden, H. Urlaub, P. Cramer, Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mol. Biol.* **27**, 668–677 (2020). doi: 10.1038/s41594-020-0437-1; pmid: 32541898
28. O. Viktorovskaya *et al*., Essential histone chaperones collaborate to regulate transcription and chromatin integrity. *Genes Dev.* **35**, 698–712 (2021). doi: 10.1101/gad.348431.121; pmid: 33888559
29. S. L. Squazzo *et al*., The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J.* **21**, 1764–1774 (2002). doi: 10.1093/emboj/21.7.1764; pmid: 11927560
30. G. A. Hartzog, T. Wada, H. Handa, F. Winston, Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369 (1998). doi: 10.1101/gad.12.3.357; pmid: 9450930
31. T. Wada *et al*., DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998). doi: 10.1101/gad.12.3.343; pmid: 9450929
32. H. Ehara *et al*., Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science* **363**, 744–747 (2019). doi: 10.1126/science.aav8912; pmid: 30733384
33. S. Li *et al*., The elongation factor Spn1 is a multi-functional chromatin binding protein. *Nucleic Acids Res.* **46**, 2321–2334 (2018). doi: 10.1093/nar/gkx1305; pmid: 29300974
34. L. McCullough, Z. Connell, C. Petersen, T. Formosa, The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in *Saccharomyces cerevisiae*. *Genetics* **201**, 1031–1045 (2015). doi: 10.1534/genetics.115.180794; pmid: 26416482
35. L. Farnung, M. Ochmann, M. Engeholm, P. Cramer, Structural basis of nucleosome transcription mediated by Chd1 and FACT. *Nat. Struct. Mol. Biol.* **28**, 382–387 (2021). doi: 10.1038/s41594-021-00578-6; pmid: 33846633
36. Y. Liu *et al*., FACT caught in the act of manipulating the nucleosome. *Nature* **577**, 426–431 (2020). doi: 10.1038/s41586-019-1820-0; pmid: 31775157
37. B. J. E. Martin, A. T. Chruscicki, L. J. Howe, Transcription promotes the interaction of the facilitates chromatin transactions (FACT) complex with nucleosomes in *Saccharomyces cerevisiae*. *Genetics* **210**, 869–881 (2018). doi: 10.1534/genetics.118.301349; pmid: 30237209
38. A. J. Hodges, L. M. Gloss, J. J. Wyrick, Residues in the nucleosome acidic patch regulate histone occupancy and are important for FACT binding in *Saccharomyces cerevisiae*. *Genetics* **206**, 1339–1348 (2017). doi: 10.1534/genetics.117.201939; pmid: 28468903
39. C. E. Cucinotta, A. E. Hildreth, B. M. McShane, M. K. Shirra, K. M. Arndt, The nucleosome acidic patch directly interacts with subunits of the Paf1 and FACT complexes and controls chromatin architecture in vivo. *Nucleic Acids Res.* **47**, 8410–8423 (2019). doi: 10.1093/nar/gkz549; pmid: 31226204
40. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997). doi: 10.1038/38444; pmid: 9305837
41. S. Ramachandran, K. Ahmad, S. Henikoff, Transcription and Remodeling Produce Asymmetrically Unwrapped Nucleosomal Intermediates. *Mol. Cell* **68**, 1038–1053.e4 (2017). doi: 10.1016/j.molcel.2017.11.015; pmid: 29225036
42. Y. Tsunaka, Y. Fujiwara, T. Oyama, S. Hirose, K. Morikawa, Integrated molecular mechanism directing nucleosome reorganization by human FACT. *Genes Dev.* **30**, 673–686 (2016). doi: 10.1101/gad.274183.115; pmid: 26966247
43. Ü. Uzun, T. Brown, H. Fischl, A. Angel, J. Mellor, Spt4 facilitates the movement of RNA polymerase II through the +2 nucleosomal barrier. *Cell Rep.* **36**, 109755 (2021). doi: 10.1016/j.celrep.2021.109755; pmid: 34592154
44. D. J. Kemble, L. L. McCullough, F. G. Whitby, T. Formosa, C. P. Hill, FACT disrupts nucleosome structure by binding H2A-H2B with conserved peptide motifs. *Mol. Cell* **60**, 294–306 (2015). doi: 10.1016/j.molcel.2015.09.008; pmid: 26455391
45. M. L. Kireeva *et al*., Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription. *Mol. Cell* **9**, 541–552 (2002). doi: 10.1016/S1097-2765(02)00472-0; pmid: 11931762
46. C. Thiriet, J. J. Hayes, Replication-independent core histone dynamics at transcriptionally active loci in vivo. *Genes Dev.* **19**, 677–682 (2005). doi: 10.1101/gad.1265205; pmid: 15769942
47. H. Kimura, P. R. Cook, Kinetics of core histones in living human cells: Little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.* **153**, 1341–1354 (2001). doi: 10.1083/jcb.153.7.1341; pmid: 11425866
48. C. Evrin *et al*., Spt5 histone binding activity preserves chromatin during transcription by RNA polymerase II. *EMBO J.* **41**, e109783 (2022). doi: 10.15252/embj.2021109783; pmid: 35102600
49. A. K. Singh, F. Mueller-Planitz, Nucleosome positioning and spacing: From mechanism to function. *J. Mol. Biol.* **433**, 166847 (2021). doi: 10.1016/j.jmb.2021.166847; pmid: 33539878
50. O. I. Kulaeva *et al*., Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat. Struct. Mol. Biol.* **16**, 1272–1278 (2009). doi: 10.1038/nsmb.1689; pmid: 19935686
51. D. A. Gaykalova *et al*., Structural analysis of nucleosomal barrier to transcription. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5787–E5795 (2015). doi: 10.1073/pnas.1508371112; pmid: 26460019
52. M. Filipovski, J. H. M. Soffers, S. M. Vos, L. Farnung, Structural basis of nucleosome retention during transcription elongation.

Science **376**, 1313–1316 (2022). doi: 10.1126/science.abo3851; pmid: 35709268
53. T. Kujirai *et al*., Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595–598 (2018). doi: 10.1126/science.aau9904; pmid: 30287617
54. A. T. Anshabo, R. Milne, S. Wang, H. Albrecht, CDK9: A comprehensive review of its biology, and its role as a potential target for anti-cancer agents. *Front. Oncol.* **11**, 678559 (2021). doi: 10.3389/fonc.2021.678559; pmid: 34041038
55. R. Dronamraju *et al*., Casein kinase II phosphorylation of Spt6 enforces transcriptional fidelity by maintaining Spn1-Spt6 interaction. *Cell Rep.* **25**, 3476–3489.e5 (2018). doi: 10.1016/j.celrep.2018.11.089; pmid: 30566871
56. K. Mayanagi *et al*., Structural visualization of key steps in nucleosome reorganization by human FACT. *Sci. Rep.* **9**, 10183 (2019). doi: 10.1038/s41598-019-46617-7; pmid: 31308435
57. E. Gouot *et al*., Casein kinase 2 mediated phosphorylation of Spt6 modulates histone dynamics and regulates spurious transcription. *Nucleic Acids Res.* **46**, 7612–7630 (2018). doi: 10.1093/nar/gky515; pmid: 29905868
58. L. G. Bedard *et al*., Quantitative analysis of dynamic protein interactions during transcription reveals a role for casein kinase II in polymerase-associated factor (PAF) complex phosphorylation and regulation of histone H2B monoubiquitylation. *J. Biol. Chem.* **291**, 13410–13420 (2016). doi: 10.1074/jbc.M116.727735; pmid: 27143358
59. H. Basnet *et al*., Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. *Nature* **516**, 267–271 (2014). doi: 10.1038/nature13736; pmid: 25252977
60. B. Kastner *et al*., GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008). doi: 10.1038/nmeth1139; pmid: 18157137
61. D. Vasudevan, E. Y. D. Chua, C. A. Davey, Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.* **403**, 1–10 (2010). doi: 10.1016/j.jmb.2010.08.039; pmid: 20800598

ACKNOWLEDGMENTS

The cryo-EM experiments were performed with the Krios G4 microscope at the RIKEN Yokohama cryo-EM facility and with the Krios G4 microscope supported by the Kurumizaka ERATO Chromatin Atlas Project (JPMJER1901). We thank M. Aoki, M. Goto, M. Henmi, T. Uejima, K. Katsura, and S. Kohno (RIKEN) for their assistance with protein preparation; T. Ito, Y. Iikura, and J. Kato (Univ. Tokyo) for their assistance with nucleosome preparation; and M. Ogasawara and Y. Takizawa (Univ. Tokyo) for helpful advice on cryo-EM data collection and analysis. **Funding:** This work was supported by the Japan Society for the Promotion of Science (KAKENHI grant JP20H03201 to H.E. and T.K.; KAKENHI grant JP20H05690 to T.K. and S.S.; KAKENHI grants JP20H00449 and JP18H05534 to H.K.); The Sumitomo Foundation (Grant for Basic Science Research Projects 190484 to T.K.); Japan Science and Technology Agency ERATO (grant JPMJER1901 to H.K.); the Platform Project for Supporting Drug Discovery and Life Science Research (BINDS) from AMED (grant JP21am0101082 to S.S. and M.S. and grant JP21am0101076 to H,K,); and the Research Support Project for Life Science and Drug Discovery (BINDS) from AMED (grant JP22ama121009 to H.K.).
Author contributions: Conceptualization: H.E., T.K., H.K., S.S.; Funding acquisition: H.E., T.K., M.S., H.K., S.S.; Investigation: H.E., T.K.; Methodology: H.E., T.K., M.S.; Project administration: H.K., S.S.; Supervision: H.K., S.S.; Visualization: H.E., T.K., H.K., S.S.; Writing – original draft: H.E., T.K., H.K., S.S.; Writing – review and editing: H.E., T.K., H.K., S.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The cryo-EM maps and coordinates have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB), respectively, with accession codes: EMD-33313 and 7XN7 (EC), EMD-33424 and 7XSE (EC42), EMD-33436 and 7XSX (EC49), EMD-33441 and 7XT7 (EC49B), EMD-33450 and 7XTI (EC58^{hex}), EMD-33447 and 7XTD (EC58^{oct}), and EMD-33437 and 7XSZ (EC115). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abp9466
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 and S2
References (*62*–*93*)
MDAR Reproducibility Checklist
Movies S1 to S3

Submitted 10 March 2022; accepted 5 August 2022
Published online 18 August 2022
10.1126/science.abp9466

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

1848 SOCIETY
AAAS

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

NEURODEVELOPMENT

Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals

Anneline Pinson, Lei Xing†, Takashi Namba†, Nereo Kalebic†, Jula Peters, Christina Eugster Oegema, Sofia Traikov, Katrin Reppe, Stephan Riesenberg, Tomislav Maricic, Razvan Derihaci, Pauline Wimberger, Svante Pääbo, Wieland B. Huttner*

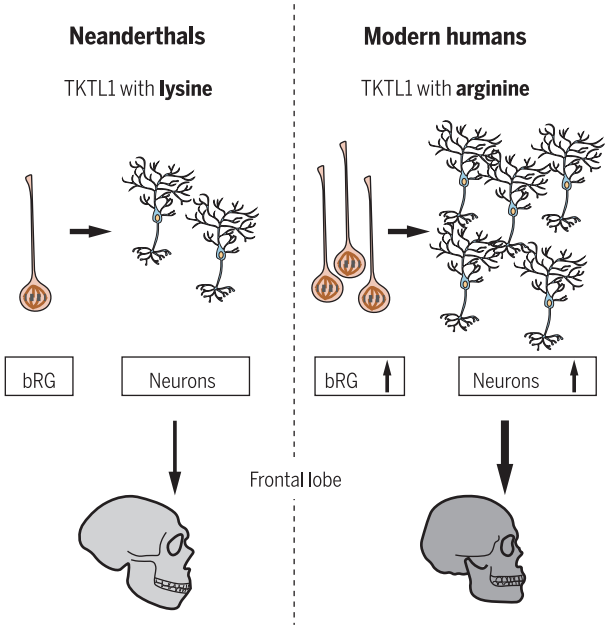
INTRODUCTION: The evolutionary expansion of the neocortex and the concomitant increase in neuron production are considered to be a basis for the increase in cognitive abilities that occurred during human evolution. Endocast analyses reveal that the endocranial volume of modern humans and Neanderthals was similar, suggesting similar brain and neocortex size. But whether similar neocortex size implies similar neocortical neuron production remains unclear.

RATIONALE: *Transketolase-like 1* (*TKTL1*) is a gene from the transketolase family that in fetal human neocortex is preferentially expressed in the two classes of neuroprogenitors, the apical progenitors in the ventricular zone and the basal progenitors in the subventricular zone. The latter class of neuroprogenitors comprises two major types, the basal intermediate progenitors (bIPs) and the basal radial glia (bRG, also called outer radial glia). bRG exhibit cellular processes that promote their ability to self-amplify, and are the neuroprogenitor type considered to be a driver of the increase in cortical neuron production, which is a hallmark of the evolution of the human neocortex. Reflecting their cell polarity, bRG undergo repeated asymmetric divisions that self-renew the bRG and generate one neuron each. Thereby, bRG generate more neurons over time than the other type of neuron-generating basal neuroprogenitors, the process-lacking bIPs whose neurogenic divisions are symmetric self-consuming. TKTL1 is one of the few proteins with a single amino acid substitution found in essentially all present-day humans but absent from extinct archaic humans, the Neanderthals and Denisovans, and other primates. This human-specific amino acid substitution in TKTL1 is a lysine in apes and archaic humans but an arginine in modern humans. We therefore investigated (i) whether TKTL1 has a role in neocortex development and affects neuroprogenitor numbers and (ii) whether both archaic TKTL1 (aTKTL1) and modern human TKTL1 (hTKTL1) exert similar effects on neuroprogenitors during neocortex development.

RESULTS: When expressed in mouse embryo neocortex, which lacks *TKTL1* expression, hTKTL1 increased the abundance of bRG without affecting that of bIPs and that of apical progenitors. The effect of TKTL1 on bRG abundance was limited to hTKTL1; aTKTL1, which differs only by one amino acid, was unable to increase bRG abundance. The greater bRG abundance upon hTKTL1 expression resulted in an increase in cortical neuron production over time, specifically of the late-born upper-layer neurons rather than of the early-born deep-layer neurons. In the folded (gyrencephalic) developing ferret neocortex, *hTKTL1* expression increased not only bRG abundance but also the proportion of bRG with multiple processes, a hallmark of bRG that can self-amplify. As a consequence of this effect, gyrus size increased.

In fetal human neocortex, hTKTL1 was essential to maintain the full number of bRG, as CRISPR-Cas9–mediated hTKTL1 knockout in fetal human neocortical tissue reduced this number. To further demonstrate the relevance of this effect, we converted hTKTL1 to the Neanderthal variant aTKTL1 in human embryonic stem cells and generated minibrain structures called cerebral organoids. The aTKTL1-expressing organoids contained fewer bRG and neurons, hence the human-specific lysine-to-arginine substitution in hTKTL1 is essential for maintaining the full number of bRG and neurons in this human brain model. In fetal human neocortex, hTKTL1 expression in neuroprogenitors increased during the course of neurogenesis and was particularly high in the developing frontal lobe as compared to the developing occipital lobe.

As to its mechanism of action, hTKTL1 increased bRG abundance via two metabolic pathways, the pentose phosphate pathway (PPP) followed by fatty acid synthesis. Inhibition of the PPP or of fatty acid synthesis, using a variety of specific inhibitors, completely suppressed the hTKTL1-induced increase in bRG abundance in embryonic mouse neocortex and reduced bRG numbers in fetal human neocortical tissue. This metabolic action of hTKTL1, but not aTKTL1, in bRG resulted in an increase in the concentration of acetyl–coenzyme A, the critical metabolite for fatty acid synthesis. Our data suggest that hTKTL1, but not aTKTL1, promotes the synthesis of membrane lipids containing a certain type of fatty acid that are required for the outgrowth of bRG processes and hence for the increase in bRG abundance.



TKTL1 and hominin cortical neurogenesis. The single lysine-to-arginine substitution in modern human TKTL1 leads to greater bRG numbers than in Neanderthals. These bRG in turn generate more neocortical neurons in modern humans. Because TKTL1 expression in fetal human neocortex is particularly high in the developing frontal lobe, these findings imply that the frontal lobe of modern humans contains more neurons than that of Neanderthals.

CONCLUSION: In light of our finding that *TKTL1* expression in fetal human neocortex is particularly high in the developing frontal lobe, our study implies that because of the single amino acid–based activity of hTKTL1, neocortical neurogenesis in modern humans was and is greater than it was in Neanderthals, in particular in the frontal lobe. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: huttner@mpi-cbg.de
†These authors contributed equally to this work.
Cite this article as A. Pinson *et al.*, *Science* **377**, eabl6422 (2022). DOI: 10.1126/science.abl6422

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl6422

RESEARCH ARTICLE

NEURODEVELOPMENT

Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals

Anneline Pinson[1], Lei Xing[1]†, Takashi Namba[1]†‡, Nereo Kalebic[1]†§, Jula Peters[1], Christina Eugster Oegema[1], Sofia Traikov[1], Katrin Reppe[1], Stephan Riesenberg[2], Tomislav Maricic[2], Razvan Derihaci[3], Pauline Wimberger[3], Svante Pääbo[2], Wieland B. Huttner[1]*

Neanderthal brains were similar in size to those of modern humans. We sought to investigate potential differences in neurogenesis during neocortex development. Modern human transketolase-like 1 (TKTL1) differs from Neanderthal TKTL1 by a lysine-to-arginine amino acid substitution. Using overexpression in developing mouse and ferret neocortex, knockout in fetal human neocortical tissue, and genome-edited cerebral organoids, we found that the modern human variant, hTKTL1, but not the Neanderthal variant, increases the abundance of basal radial glia (bRG) but not that of intermediate progenitors (bIPs). bRG generate more neocortical neurons than bIPs. The hTKTL1 effect requires the pentose phosphate pathway and fatty acid synthesis. Inhibition of these metabolic pathways reduces bRG abundance in fetal human neocortical tissue. Our data suggest that neocortical neurogenesis in modern humans differs from that in Neanderthals.

Whether and how cognitive abilities of modern humans might differ from those of extinct archaic humans such as Neanderthals remains a matter of debate (*1*). Discovery of artifacts and art of presumably Neanderthal derivation (*2*, *3*) fuels the debate. Cognitive abilities reside primarily in the neocortex, the largest and most recently evolved part of the brain, which is present only in mammals. The evolutionary expansion of the neocortex and the concomitant increase in neuron production are considered to be a basis for the increase in cognitive abilities that occurred during human evolution (*4*). Analyses of endocasts indicate that the endocranial volume of modern humans and Neanderthals was similar, suggesting similar brain and neocortex size (*5*). But whether similar neocortex size implies similar neocortical neuron production has remained unclear.

One approach to address this question would be to compare key features of neocortex development between modern humans and Neanderthals, using appropriate model systems and focusing on the actions of modern human versus Neanderthal variants of key genes that govern neocortex development. Of particular interest here are genes that influence the behavior of neural progenitor cells (neuroprogenitors) in the fetal neocortex, as their abundance and proliferative capacity determine the number of cortical neurons generated during development (*6*, *7*).

Two principal classes of neuroprogenitors are present in the developing neocortex, referred to as apical progenitors (APs) and basal progenitors (BPs). APs, the primary class, reside in the ventricular zone (VZ). After the onset of neurogenesis, the major AP type is the apical (or ventricular) radial glia (aRG). Rather than producing neurons, aRG generate mostly BPs, the secondary class of neuroprogenitors. Newborn BPs migrate to the subventricular zone (SVZ), from where they generate most of the cortical neurons (*7*).

Two types of BPs have been characterized, referred to as basal intermediate progenitors (bIPs) and basal (or outer) radial glia (bRG). In mammals with a small and lissencephalic neocortex, such as mice, ~90% of the BPs are neurogenic bIPs that typically divide once to give rise to two neurons [(*7*) and references therein]. bRG constitute only ~10% of the BPs in embryonic mouse neocortex (*8*–*10*). In contrast, in ferrets (a gyrencephalic carnivore) and in primates such as marmosets, macaques, and humans, bRG constitute ~50% of the BP pool (*8*, *11*–*13*). This increase in the relative abundance of bRG is considered to drive the increase in cortical neuron production that is a hallmark of the evolution of the human neocortex (*6*, *14*, *15*).

In contrast to bIPs, neurogenic divisions of bRG are typically asymmetric, generating a bRG (self-renewal) and a neuron (*7*, *12*, *13*). This mode of cell division reflects the presence of cell polarity of bRG, evident as a basal process, which is lacking in bIPs (*7*, *12*, *13*). bRG generate more cortical neurons over time than bIPs (*12*).

A gene that could influence the behavior of neuroprogenitors in fetal human neocortex is *transketolase-like 1* (*TKTL1*) (*16*). TKTL1 belongs to the transketolase family of enzymes. It operates in the pentose phosphate pathway (PPP), a metabolic pathway linked to glycolysis (*17*–*19*). We focused on *TKTL1* because: (i) *TKTL1* is preferentially expressed in neuroprogenitors, including bRG, of fetal human neocortex (*20*–*23*). (ii) *TKTL1*, implicated in human tumors (*24*, *25*) and tumor cell proliferation (*19*, *26*), may also increase neuroprogenitor numbers. (iii) TKTL1 is one of the few proteins with an amino acid substitution found in essentially all present-day humans and absent from extinct archaic humans, Neanderthals and Denisovans, and other primates (*27*).

The human-specific amino acid substitution at residue 317 in the long isoform of TKTL1 (corresponding to amino acid residue 261 in the short isoform) is a lysine in apes and archaic humans but an arginine in modern humans (*27*). We therefore investigated (i) whether TKTL1 has a role in neocortex development and affects neuroprogenitor behavior and (ii) whether both archaic TKTL1 (aTKTL1) and modern human TKTL1 (hTKTL1) exert similar effects on neuroprogenitors during neocortex development. We find that hTKTL1 but not aTKTL1 increases the abundance of bRG, but not bIPs, in developing neocortex. We suggest that more neocortical neurons are generated during neurogenesis in modern humans than in Neanderthals.

TKTL1 expression in fetal human neocortex

Analysis of published transcriptome datasets (*21*–*23*) revealed that in the human neocortex at 13 to 16 postconception weeks (PCW), *TKTL1* mRNA is expressed in the VZ and inner and outer SVZ (iSVZ and oSVZ) but not the cortical plate (CP) (fig. S1A). Accordingly, *TKTL1* mRNA is expressed in aRG and bRG isolated from PCW 13 or PCW 16–17 human neocortex. *Tktl1* mRNA is not expressed in embryonic day 14.5 (E14.5) mouse neocortex (fig. S1A). Further analysis showed that the *TKTL1* isoform expressed in fetal human neocortex is the short isoform, encoding a 540–amino acid protein (fig. S1, B, D, and E), with the amino acid change at position 261 (fig. S1H).

Highest *TKTL1* expression in frontal lobe

TKTL1 expression in fetal human neocortex by reverse transcription quantitative polymerase chain reaction was found to increase with development (PCW 9 to 15; fig. S1, D and E). Consistent with a previous report (*28*), *TKTL1* expression increased at PCW 17 in the frontal lobe, but not the occipital lobe, of fetal human neocortex (fig. S1, F and G). This increased expression in frontal lobe was already detectable in VZ and SVZ of PCW 11 neocortex by in situ hybridization (fig. S1C).

[1]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [2]Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [3]Technische Universität Dresden, Universitätsklinikum Carl Gustav Carus, Klinik und Poliklinik für Frauenheilkunde und Geburtshilfe, 01307 Dresden, Germany.
***Corresponding author. Email: huttner@mpi-cbg.de** †These authors contributed equally to this work. ‡Present address: Neuroscience Center, HiLIFE–Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. §Present address: Human Technopole, Milan, Italy.

hTKTL1, not aTKTL1, increases bRG abundance

As *Tktl1* mRNA is not expressed in the embryonic mouse neocortex, this system facilitated investigation of whether ectopic expression of *hTKTL1* or *aTKTL1* would affect neuroprogenitor abundance. For this purpose, we performed in utero electroporation (IUE) of a plasmid encoding hTKTL1 or aTKTL1, or of an empty vector (control), along with a plasmid encoding green fluorescent protein (GFP) to allow the identification of the progeny of the electroporated cells, in mouse lateral neocortex at E13.5, followed by analysis at E15.5. Immunofluorescence confirmed expression of hTKTL1 and aTKTL1 in electroporated neocortex (fig. S2A).

Immunofluorescence for the mitotic markers phospho-vimentin (pVim) (Fig. 1A) and phospho-histone H3 (PH3) (fig. S2C) revealed no change in the abundance of GFP$^+$ ventricular mitoses, which correspond to mitotic aRG, upon hTKTL1 or aTKTL1 IUE (Fig. 1B and fig. S2D). In contrast, the abundance of GFP$^+$ mitotic BPs was increased in hTKTL1-electroporated, but not in aTKTL1-electroporated, neocortex (Fig. 1C and fig. S2E). Likewise, immunofluorescence for the cycling cell marker PCNA (fig. S2F) revealed no change in the percentage of cycling APs (fig. S2G) but an increase in the percentage of cycling BPs upon hTKTL1, but not aTKTL1, electroporation (fig. S2H). We next assessed the presence (bRG) versus absence (bIPs) of pVim$^+$ processes among the GFP$^+$ BPs (*12*, *13*). Whereas the abundance of mitotic bIPs was not affected (Fig. 1D), the abundance of process-bearing pVim$^+$ GFP$^+$ BPs (i.e., bRG) was increased by a factor of 4 upon hTKTL1, but not aTKTL1, expression (Fig. 1E). Thus, mitotic bRG accounted for only ~10% of all mitotic BPs in control [as shown previously (*8*–*10*)] and in aTKTL1-electroporated embryonic mouse neocortex, but accounted for 33% of all mitotic BPs in hTKTL1-electroporated neocortex (fig. S2B). Hence, hTKTL1, but not aTKTL1, increases selectively bRG abundance without affecting bIP abundance.

In embryonic mouse medial neocortex, bRG abundance is greater than in lateral neocortex and is similar to that in developing gyrencephalic neocortex, with human bRG-like gene expression (*29*). Similar to embryonic mouse lateral neocortex, *hTKTL1* expression in medial neocortex increased mitotic bRG (fig. S2, I, K, and L), but not mitotic bIP (fig. S2J), abundance.

hTKTL1, not aTKTL1, increases Sox2$^+$/Tbr2$^-$ bRG

E15.5 mouse lateral neocortex subjected to either hTKTL1, aTKTL1, or control IUE at E13.5 was analyzed by immunofluorescence for Sox2, which in the SVZ is a marker of proliferative BPs, and Tbr2, a marker of neurogenic BPs (fig. S3A). Compared to control, *hTKTL1*- and *aTKTL1*-electroporated neocortex showed no change in the percentage of GFP$^+$ cells in the VZ that were Tbr2$^+$ (fig. S3B), Sox2$^+$ (fig. S3D), or Sox2$^+$ and Tbr2$^-$ (fig. S3F). In contrast, hTKTL1, but not aTKTL1, increased the percentage of GFP$^+$ cells in the SVZ that were Sox2$^+$ (fig. S3E), without affecting the percentage of GFP$^+$ cells in the SVZ that were Tbr2$^+$ (fig. S3C). The Sox2$^+$ BPs increased by hTKTL1 were Tbr2$^-$ (i.e., bRG) (fig. S3, G and H). We conclude that the increase in bRG abundance induced by expression of hTKTL1, but not aTKTL1, involves an increase in bRG proliferation.

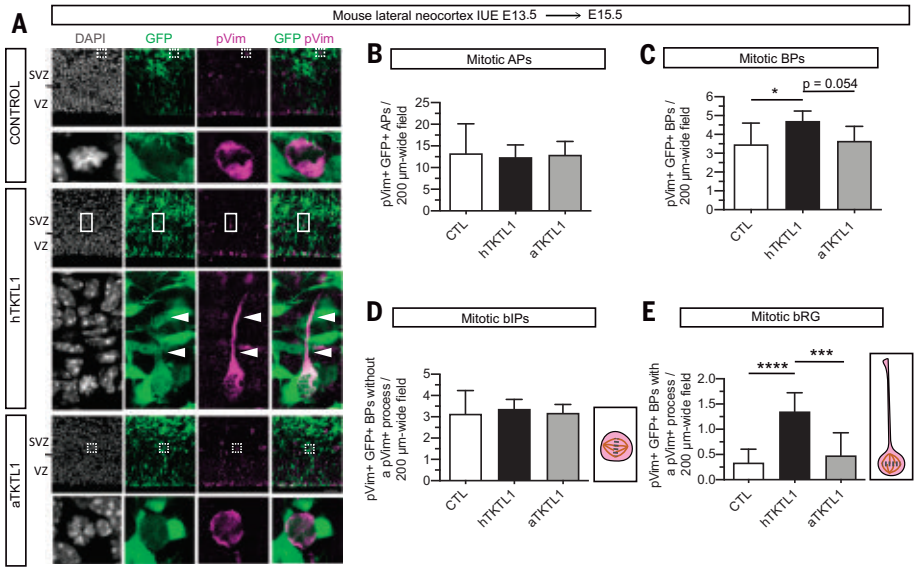


Fig. 1. Modern human TKTL1, but not archaic TKTL1, when expressed in embryonic mouse neocortex, increases mitotic bRG abundance. Mouse neocortex E13.5 IUE with GFP plasmid, together with either empty (control, CTL), hTKTL1, or aTKTL1 plasmid; analyses: E15.5. (**A**) GFP/pVim (green/magenta) immunofluorescence plus DAPI staining (gray). Bottom rows, white boxed areas at higher magnification; dashed boxed areas: GFP$^+$/pVim$^+$/BP without pVim$^+$ process (mitotic bIP); solid boxed area: GFP$^+$/pVim$^+$/BP with pVim$^+$ basal process (arrowheads) (mitotic bRG). Scale bar, 40 µm. (**B** to **E**) Quantifications in 200-µm-wide fields. Means of 8 embryos. Error bars, SD. (B) pVim$^+$ GFP$^+$ mitotic APs. One-way analysis of variance (ANOVA). (C) Total pVim$^+$/GFP$^+$/BPs. One-way ANOVA with Tukey post hoc test, *P < 0.05. (D) pVim$^+$/GFP$^+$/BPs without pVim$^+$ process (mitotic bIPs). One-way ANOVA. (E) pVim$^+$/GFP$^+$/BPs with pVim$^+$ process (mitotic bRG). One-way ANOVA with Tukey post hoc test, ****P < 0.0001, ***P < 0.001. Here and in the remaining figures, data reported without P values are not significant.

bRG generate more neurons than bIPs

We considered the implications of the selective increase in bRG abundance upon hTKTL1 expression for cortical neurogenesis, based on a previous mathematical modeling of cortical neuron production by neurogenic divisions of bIPs versus bRG (*12*). In this model, as summarized in Fig. 2, A and B, both bIPs and bRG are generated by repeated asymmetric self-renewing divisions of aRG. Neuron-generating divisions of bIPs are typically symmetric consumptive, with one bIP giving rise to two neurons (Fig. 2A). In contrast, reflecting the cell polarity of bRG, in particular the presence of a basal process, neuron-generating divisions of bRG are typically asymmetric self-renewing, with one bRG producing one neuron and one bRG (Fig. 2B) (*7*, *9*, *10*, *12*, *13*). These two consecutive, asymmetric self-renewing divisions (aRG → bRG → neuron) result in an initially delayed, but eventually much more efficient, production of cortical neurons when compared to the aRG → bIP → neuron lineage (compare panels A and B of Fig. 2). We therefore modeled the consequences of the hTKTL1-induced increase in bRG abundance for cortical neuron production, assuming one round of symmetric proliferative bRG division prior to the start of the asymmetric self-renewing neuron-generating bRG divisions (Fig. 2C), as is suggested by the hTKTL1-induced increase in Sox2$^+$ Tbr2$^-$ bRG. This resulted in an even greater increase in cortical neuron production over time (Fig. 2C). We conclude that the increase in bRG abundance by hTKTL1 but not aTKTL1 implies an increase in production of neurons over time in the developing neocortex of modern humans as compared to archaic humans.

hTKTL1 increases abundance of late-born neurons

We therefore investigated whether hTKTL1 IUE into E13.5 mouse neocortex resulted in an increase specifically in late-born neurons

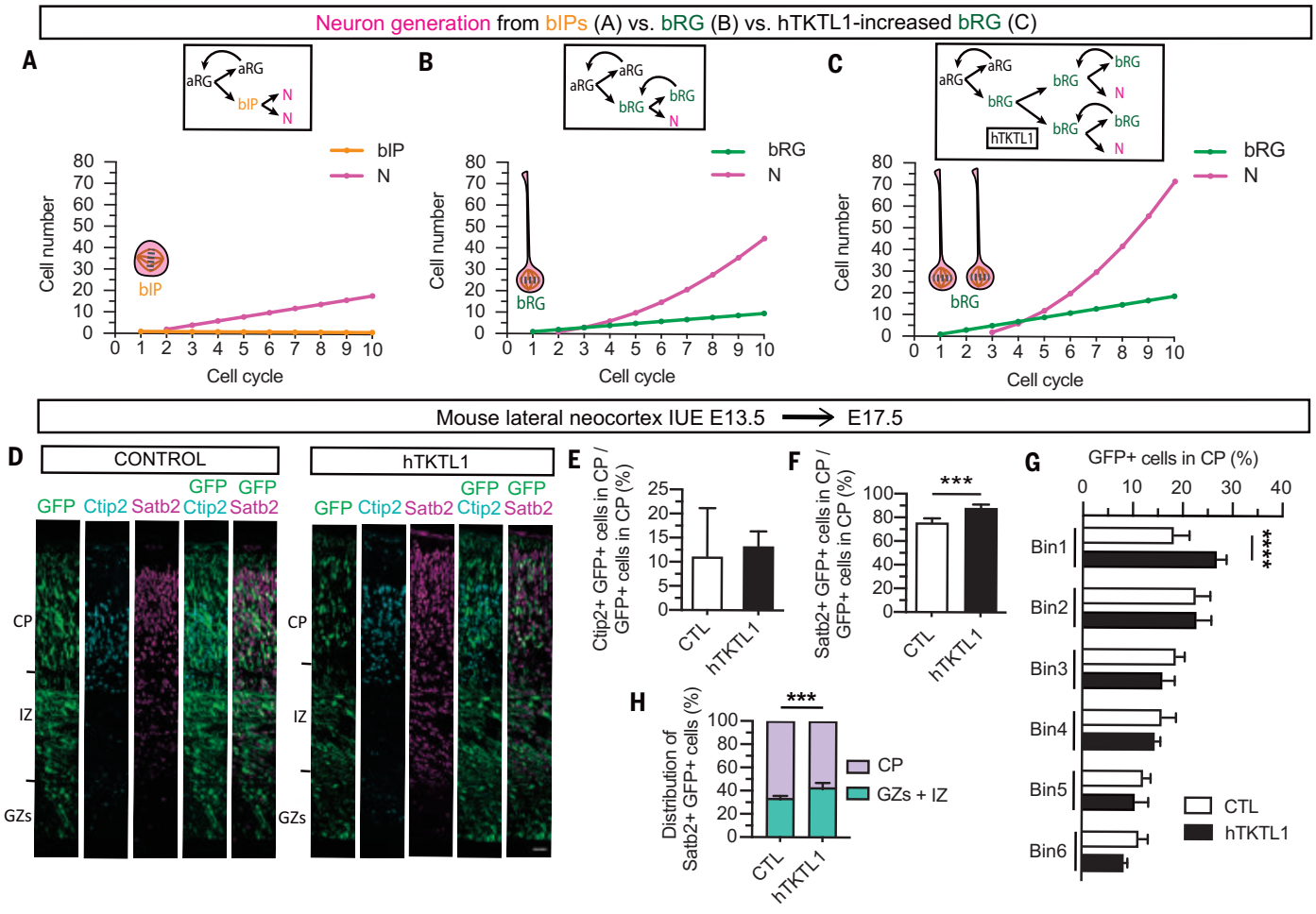


Fig. 2. The hTKTL1-induced increase in bRG in embryonic mouse neocortex results in increased production of cortical neurons, notably upper-layer neurons, at late neurogenesis. (**A** to **C**) Modeling of the number of neurons generated, over 10 cell cycles, by the aRG → bIP → neuron lineage (A), the aRG → bRG → neuron lineage (B), or the aRG → bRG → neuron lineage with one round of hTKTL1-induced symmetric proliferative bRG division (C). Curved arrows denote self-renewal. (A) and (B) are adapted from figure S4 of (*12*). (**D** to **H**) Mouse neocortex E13.5 IUE with GFP plasmid, together with either empty (control, CTL) or hTKTL1 plasmid; analyses: E17.5. (E) to (H) are means of 5 embryos. Error bars, SD. (D) GFP/Ctip2/Satb2 (green/cyan/magenta) immunofluorescence. Scale bar, 30 μm. [(E) and (F)] Percentages of GFP^+ cells in CP that are $Ctip2^+$ (E) and $Satb2^+$ (F). Unpaired Student's *t* test, (F) $^{***}P < 0.001$. (G) Percentages of the GFP^+ cells in CP that are in bins 1 to 6 (bin 1, uppermost layer; bin 6, deepest layer) of CP. Two-way ANOVA with Bonferroni post hoc test, $^{****}P < 0.0001$. (H) Distribution of $GFP^+/Satb2^+$ cells in germinal zones (GZs) plus intermediate zone (IZ) versus CP. Student's *t* test, $^{***}P < 0.001$.

at E17.5. To this end, we performed immunofluorescence for Ctip2, a marker of early-born deep-layer neurons, and for Satb2, a marker of late-born upper-layer neurons (Fig. 2D and fig. S3I). hTKTL1 expression did not affect the percentage of GFP^+ cells in the CP that expressed Ctip2 (Fig. 2E) but increased the percentage of GFP^+ cells in the CP that expressed Satb2 (Fig. 2F). Consistent with this, hTKTL1 expression increased the proportion of the GFP^+ cells in the CP that were located in the basal-most bin (Fig. 2G and fig. S3I). In addition, the analysis of the distribution of the $Satb2^+$ GFP^+ neurons across the E17.5 cortical wall revealed that in the control 66% of these neurons had reached the CP during the 4-day interval between electroporation and analysis, whereas only 57% of these neurons had reached the CP in the hTKTL1-electroporated neocortex (Fig. 2H), as would be expected if late-born neuron generation was increased in this condition. We conclude that upon hTKTL1 expression, consistent with the mathematical modeling of the consequences of a bRG increase for cortical neuron production (Fig. 2C), the production of late-born neurons, but not that of early-born neurons, is increased.

hTKTL1 increases bRG in ferret neocortex

We investigated whether hTKTL1 expression is also able to increase bRG abundance in a gyrencephalic species, the ferret, which endogenously expresses an archaic version of TKTL1 with a lysine residue instead of the arginine residue present in hTKTL1. Similar to humans, the ferret SVZ is divided into an iSVZ and an oSVZ. We performed IUE of hTKTL1 in ferret neocortex at E33, which corresponds to the start of the generation of the oSVZ and of the production of upper-layer neurons (*30*). This led to a factor of 4 increase in the abundance of mitotic bRG, but not mitotic bIPs, in iSVZ and oSVZ at postnatal day 2 (P2) (fig. S4, A to D). This bRG increase was accompanied by an increase in $PCNA^+$ cells (fig. S4, E and F) and $Sox2^+/Tbr2^-$ cells and a decrease in $Tbr2^+$ cells, in iSVZ and oSVZ (fig. S5). hTKTL1 increased the proportion of radial BPs (bRG) compared to multipolar BPs (bIPs) (fig. S4G), and within the former of bRG with multiple radial processes (*31*) at the expense of the bRG with one basal process (Fig. 3A). We conclude that ectopic expression of hTKTL1 in developing neocortex increases bRG abundance in both lissencephalic and gyrencephalic species.

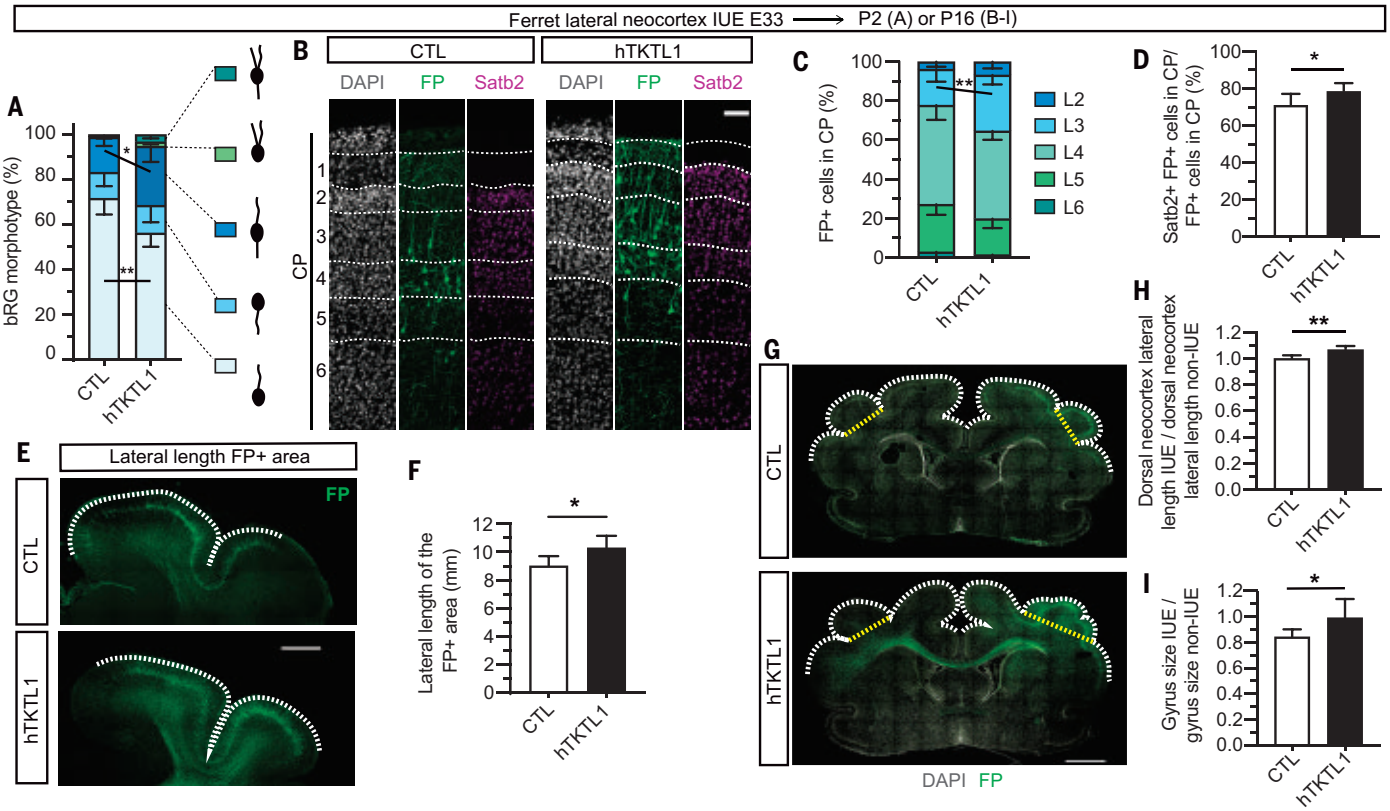


Fig. 3. Modern human TKTL1 expressed in developing ferret neocortex increases upper-layer neuron production, neocortical surface area, and gyrus size. Ferret neocortex E33 IUE with fluorescent protein (FP) plasmid, together with either empty (control, CTL) or hTKTL1 plasmid; analyses: P2 (A) or P16 [(B) to (I)]. Means of 3 (A) or 6 [(B) to (I)] embryos. Error bars, SD. (**A**) Percentages of the 5 morphotypes of PCNA$^+$ radial cells in SVZ (bRG). Two-way ANOVA with Bonferroni post hoc test, $^*P < 0.05$, $^{**}P < 0.01$. (**B**) FP/Satb2 (green/magenta) immunofluorescence plus DAPI staining (gray). White dashed lines: neuronal layers (L) 1-6 of CP. Scale bar, 75 μm. (**C**) Distribution of FP$^+$ cells across L2 to L6 of the CP. Two-way ANOVA with Bonferroni post hoc test, $^{**}P < 0.01$. (**D**) Percentage of FP$^+$ cells in CP that are Satb2$^+$. Unpaired Student's *t* test, $^*P < 0.05$. (**E**) FP (green) immunofluorescence. White dashed lines: lateral length electroporated area. Scale bar, 1 mm. (**F**) Lateral length FP$^+$ area. Unpaired Student's *t* test, $^*P < 0.05$. (**G**) FP (green) immunofluorescence plus DAPI staining (gray). White dashed lines: lateral length dorsal neocortex. Yellow dashed lines: areas measured for gyrus size. Scale bar, 2 mm. (**H**) Lateral length dorsal neocortex. (**I**) Size of the gyrus showing the greatest FP immunoreactivity. [(H) and (I)] Data expressed as ratio of electroporated (IUE) over contralateral (non-IUE) hemisphere. Unpaired Student's *t* test, (H) $^{**}P < 0.01$, (I) $^*P < 0.05$.

hTKTL1 expands ferret upper-layer neurons and neocortex

Following IUE at E33, we analyzed ferret neocortex at P16, a developmental stage when in ferrets the cortical folds have already formed (*32*, *33*). Upon hTKTL1 expression, a greater proportion of the targeted neuroprogenitor-derived progeny was now located in the upper layers of the CP, and in particular in layer 3 (Fig. 3, B and C). Corroborating this finding, hTKTL1 increased the proportion of this progeny that expressed Satb2, a transcription factor present in the majority of upper-layer neurons (Fig. 3D).

We explored whether the hTKTL1-induced increase in developing ferret neocortex in bRG abundance at P2 and upper-layer neurons at P16 affected neocortical morphology. We found (i) an increase in the lateral spread of the targeted neuroprogenitor-derived progeny at P16 (Fig. 3, E and F) and (ii) an increase in neocortical surface area as revealed by the lateral length of the dorsal neocortex (Fig. 3, G and H). Moreover, hTKTL1 expression elicited an increase in the size of the electroporated gyrus (Fig. 3, G and I). In half of the hTKTL1-electroporated neocortices (3/6), we observed a difference in the morphology of the gyral-sulcal pattern of the neocortex (fig. S6C), which, however, did not result in a significant change in the local gyrification index (fig. S6, A and B). Nonetheless, these findings demonstrate that hTKTL1 expression is sufficient to alter the morphology of a gyrencephalic neocortex.

hTKTL1 knockout reduces human bRG abundance

We examined whether, in fetal human neocortex, the endogenously expressed hTKTL1 is essential for maintaining bRG abundance. To this end, we performed a CRISPR-Cas9–mediated *hTKTL1* knockout (KO) in PCW 8–14 human neocortical tissue ex vivo (see fig. S7H for the efficiency of gRNAs). *hTKTL1* KO reduced the percentage of basal GFP$^+$ cells that were cycling progenitors (PCNA$^+$) (Fig. 4, A and B). Immunofluorescence for HOPX [which, combined with morphological analysis, can be used to identify bRG (*20*)] revealed that *hTKTL1* KO decreased the percentage of HOPX$^+$ cells that have a radial morphology (i.e., bRG) among the basal GFP$^+$ cells (Fig. 4, A and C). Hence, hTKTL1 is essential to maintain the full level of bRG during fetal human neocortical development.

aTKTL1-expressing organoids: Fewer bRG and neurons

We next explored the relevance for humans of our finding that hTKTL1, but not aTKTL1, increases bRG abundance. To this end, we made use of CRISPR-Cas9–mediated genome editing and "Neanderthalized" *TKTL1* [by converting Arg^{261} (fig. S1H) to Lys] in H9 human embryonic stem cells (ESCs). Two mock-edited

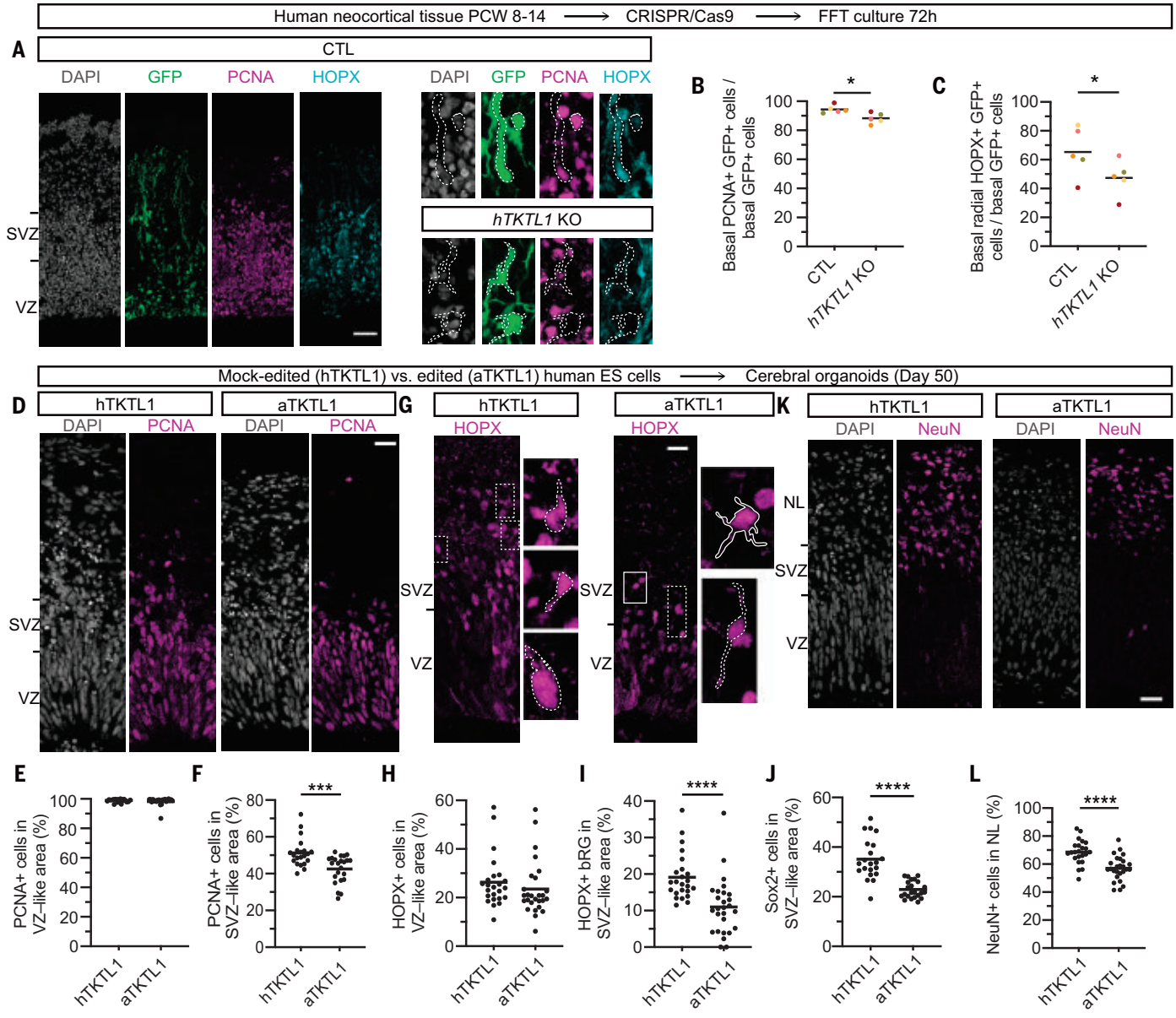


Fig. 4. *TKTL1* KO in fetal human neocortical tissue and "Neanderthalized" TKTL1 in human ESC–derived cerebral organoids reveal that modern human TKTL1 is essential to maintain the full level of bRG and neurons. (**A** to **C**) CRISPR-Cas9–mediated disruption of *hTKTL1* expression in PCW 8 to 14 fetal human neocortical tissue. Ex vivo electroporation with GFP plasmid plus complexes of recombinant Cas9 protein and gRNAs targeting *LacZ* (control, CTL) or *hTKTL1* (*hTKTL1* KO), followed by 72 hours free-floating tissue (FFT) culture. (A) GFP/PCNA/HOPX (green/magenta/cyan) immunofluorescence plus DAPI staining (gray). Left: CTL electroporation. Right: abventricular cells for CTL (top) and *hTKTL1* KO (bottom). White dashed lines: cell morphology. CTL: $GFP^+/PCNA^+/HOPX^+$ radial cell (bRG), $GFP^+/PCNA^+/$ weakly $HOPX^+$ cell (multipolar in different optical section, bIP). *hTKTL1*-KO: multipolar $GFP^+/PCNA^+/HOPX^-$ cell (bIP, top cell), multipolar $GFP^+/PCNA^-/HOPX^-$ cell (neuron, bottom cell). Scale bar, 50 μm. [(B) and (C)] Percentages of basal GFP^+ cells that are $PCNA^+$ (B) and $HOPX^+$ with radial morphology (C). Means of 5 different fetal samples. Paired Student's *t* test, $^*P < 0.05$. (**D** to **L**) Human embryonic stem cells (ESCs, H9 line) were CRISPR-Cas9–mediated genome-edited to convert *hTKTL1* (Arg) to *aTKTL1* (Lys). Organoids grown from two mock-edited (hTKTL1-1, hTKTL1-2) and two edited (aTKTL1-1, aTKTL1-2) lines; analyses: day 50. (D) PCNA (magenta) immunofluorescence plus DAPI-staining (gray). Scale bar, 25 μm. [(E) and (F)] Percentages of $PCNA^+$ cells in VZ-like (E) and SVZ-like (F) areas. Means of 21 hTKTL1 (9 hTKTL1-1, 12 hTKTL1-2) and 23 aTKTL1 (12 aTKTL1-1, 11 aTKTL1-2) organoids. Mann-Whitney U test, $^{***}P < 0.001$. (G) HOPX (magenta) immunofluorescence. Right sides: white boxes at higher magnification; dashed lines: radial cells; solid line: multipolar cell. Scale bar, 25 μm. [(H) and (I)] Percentages of $HOPX^+$ cells in VZ-like area (H) and $HOPX^+$ radial cells in SVZ-like area (I). Means of 24 hTKTL1 (11 hTKTL1-1, 13 hTKTL1-2) and 27 aTKTL1 (19 aTKTL1-1, 8 aTKTL1-2) organoids. Mann-Whitney U test, (I) $^{****}P < 0.0001$. (J) Percentages of $Sox2^+$ cells in the VZ-like area. Means of 21 hTKTL1 (9 hTKTL1-1, 12 hTKTL1-2) and 23 aTKTL1 (12 aTKTL1-1, 11 aTKTL1-2) organoids. Mann-Whitney U test, $^{****}P < 0.0001$. (K) NeuN (magenta) immunofluorescence plus DAPI-staining (gray). Scale bar, 25 μm. (L) Percentage of $NeuN^+$ cells in neuronal layer (NL). Means of 23 hTKTL1 (9 hTKTL1-1, 14 hTKTL1-2) and 26 aTKTL1 (12 aTKTL1-1, 14 aTKTL1-2) organoids. Unpaired Student's *t* test, $^{****}P < 0.0001$.

hTKTL1 (Arg) and two edited *aTKTL1* (Lys) H9 lines were generated, all of which (i) expressed established markers of pluripotency (fig. S7, A and B); (ii) did not show any chromosomal aneuploidies and large-scale chromosomal duplications and deletions in comparison to the mother H9 ESC line used for genome editing, as revealed by shallow DNA sequencing (fig. S7C); and (iii) showed identical karyograms (fig. S7D). The two mock-edited and the two *aTKTL1*-edited H9 lines were used to generate cerebral organoids (fig. S8A) that were analyzed at day 50. In situ hybridization showed that *TKTL1* mRNA is expressed in the germinal zones of the cerebral organoids (fig. S8B).

The percentage of cycling cells ($PCNA^+$) in the VZ-like area did not differ between the hTKTL1-expressing and aTKTL1-expressing organoids (Fig. 4, D and E). However, the percentage of cycling cells in the SVZ-like area was reduced in the aTKTL1-expressing organoids (Fig. 4, D and F). Moreover, aTKTL1-expressing organoids showed no change in the percentage of $HOPX^+$ cells (Fig. 4, G and H) or of $Sox2^+$ cells (fig. S8, C and D) in the VZ-like area, but a decrease in the percentage of $HOPX^+$ cells with a radial morphology (bRG) (Fig. 4, G and I) and of $Sox2^+$ cells (Fig. 4J and fig. S8C) in the SVZ-like area. Consistent with the reduction in bRG, the *TKTL1*-"Neanderthalized" cerebral organoids exhibited a reduction in the percentage of cells in the layer (referred to as neuronal layer) that express the neuronal markers NeuN (Fig. 4, K and L) and Hu (fig. S8, E and F). The reduction in bRG and neurons observed in aTKTL1-expressing organoids was not due to an increase in cell death (fig. S8G). Taken together, these data show that the Arg^{261} in hTKTL1 is essential for the generation of the full level of bRG and neurons in human cerebral organoids.

Pentose phosphate pathway inhibition abolishes hTKTL1-induced bRG increase

We sought to obtain insight into the mechanism of action of hTKTL1. TKTL1 belongs to the transketolase family (*18*), and TKT, the founding member of this family, is known to operate in the pentose phosphate pathway (PPP), an action also reported for TKTL1 (*18*, *19*) (Fig. 5A). Thus, TKTL1 has been reported to cleave xylulose 5-phosphate, a metabolite in the PPP, into glyceraldehyde 3-phosphate and acetate (*19*) (Fig. 5A). We therefore examined a possible role of the PPP in the hTKTL1-induced increase in bRG abundance. We used 6-aminonicotinamide (6-AN), an NADP analog, to inhibit 6-phosphogluconate dehydrogenase (6PGDH), the enzyme converting 6-phosphogluconate into ribulose 5-phosphate in the PPP (Fig. 5A). 6-AN is 400 times more potent to inhibit 6PGDH than to inhibit other NADP-dependent enzymes (*34*). E13.5 mouse neocortex was electroporated with a plasmid encoding hTKTL1 or an empty vector, along with a plasmid encoding GFP. The brain was dissected 24 hours after IUE and subjected to HERO culture for 24 hours in the absence or presence of 6-AN (fig. S9A). 6-AN at 50 and 100 μM completely suppressed the hTKTL1-induced increase in mitotic bRG (Fig. 5C and fig. S9, B and C) without significantly affecting the essentially equal levels upon control versus hTKTL1 IUE of mitotic bIPs (Fig. 5B) and mitotic aRG (fig. S9D). The lack of effect on bIP and aRG numbers by 6-AN treatment implies that this drug, even at 100 μM, did not exert unspecific effects on neuroprogenitors. Consistent with these data for mitotic neuroprogenitors, treatment with 50 μM or 100 μM 6-AN completely suppressed the hTKTL1-induced increase in the percentage of GFP^+ cells in the SVZ that were $Sox2^+$ (fig. S9, E and G) but had no effect on the percentage of these cells in the VZ (fig. S9F), nor did it affect the percentage of GFP^+ cells in the SVZ that were $Sox2^+$ upon control IUE (fig. S9G).

To corroborate these results, we inhibited 6PGDH with a different pharmacological inhibitor, 1-hydroxy-8-methoxyanthracene-9,10-dione (S3). Using the same experimental model, we found that the hTKTL1-induced increase in mitotic bRG was partially reduced upon treatment with 10 μM S3 and abolished upon treatment with 20 μM S3 (fig. S10, A and D). This treatment had no effect on mitotic aRG (fig. S10B) and mitotic bIP abundance (fig. S10C). Accordingly, treatment with 10 or 20 μM S3 suppressed the hTKTL1-induced increase in the percentage of GFP^+ cells in the SVZ that are $Sox2^+$ (fig. S10, E and G) without affecting the percentage of these cells in the VZ (fig. S10F). Taken together, these results demonstrate that the hTKTL1-induced increase in bRG abundance requires the PPP, specifically the 6-phosphogluconate to ribulose 5-phosphate conversion (see Fig. 5A).

PPP inhibition reduces human neocortical bRG

To examine the physiological relevance of these data, we subjected PCW 11–13 human neocortical tissue, which endogenously expresses *hTKTL1*, to free-floating tissue (FFT) culture for 2 days in the presence or absence of the PPP inhibitors 6-AN or S3. Treatment with 50 or 100 μM 6-AN reduced the abundance of mitotic bRG by ~60% (Fig. 5, D and F) but did not affect the abundance of mitotic bIPs (Fig. 5, D and E). Likewise, treatment with 5, 10, or 20 μM S3 reduced the abundance of mitotic bRG by ~40 to 55% (fig. S10, H and J) but did not affect the abundance of mitotic bIPs (fig. S10I). The reductions in bRG observed upon 6-AN and S3 treatments were not due to an increase in apoptosis (fig. S9, H and I, and fig. S10, K and L). We conclude that the 6-phosphogluconate to ribulose 5-phosphate conversion step in the PPP, which is required for the hTKTL1-induced increase in bRG abundance in embryonic mouse neocortex (Fig. 5C; fig. S9, C and G; and fig. S10, D and G), is also required for the maintenance of the physiological bRG level in fetal human neocortex, indicating a role of the PPP in this maintenance.

Fatty acid synthesis inhibition abolishes hTKTL1-induced bRG increase

One of the metabolites generated by the reactions in the PPP downstream of the 6-phosphogluconate to ribulose 5-phosphate conversion step is glyceraldehyde 3-phosphate, which via conversion to pyruvate in the glycolysis pathway can eventually give rise to acetyl–coenzyme A (CoA) (Fig. 5A). In addition, TKTL1 has been reported to cleave xylulose 5-phosphate, a metabolite in the PPP, into glyceraldehyde 3-phosphate and acetate. The latter, too, can be further processed to acetyl-CoA (*19*) (Fig. 6A). As it has been reported that TKTL1—via elevated acetyl-CoA levels—promotes fatty acid synthesis (*19*), we explored whether increased fatty acid synthesis is involved in the hTKTL1-induced increase in bRG abundance. Fatty acid synthesis starts with the conversion of acetyl-CoA to malonyl-CoA by the enzyme acetyl-CoA carboxylase (ACACA) (Fig. 6A). We therefore subjected mouse neocortex, hTKTL1- and control-electroporated at E13.5, to HERO culture starting at E14.5 in the absence or presence of 5-(tetradecyloxy)-2-furoic acid (TOFA), an inhibitor of ACACA. Analysis after 24 hours of HERO culture revealed that the hTKTL1-induced increase in mitotic bRG was halved upon treatment with 10 μM, and abolished with 20 or 50 μM, TOFA (Fig. 6C and fig. S11, A and C). TOFA treatment did not affect the equal levels upon control versus hTKTL1 IUE of mitotic APs (fig. S11B) or mitotic bIPs (Fig. 6B). Consistent with these data for mitotic neuroprogenitors, TOFA treatment suppressed the hTKTL1-induced increase in the percentage of GFP^+ cells in the SVZ that were $Sox2^+$ (fig. S11, D and F) but had no effect on the percentage of these cells in the VZ (fig. S11E), nor did it affect the percentage of GFP^+ cells in the SVZ that were $Sox2^+$ upon control IUE (fig. S11F).

These results were corroborated using a second inhibitor of ACACA, ND-646. Treatment with 5 μM ND-646, using the same experimental model, abolished the hTKTL1-induced increase in the percentage of GFP^+ cells in the SVZ that were $Sox2^+$ (fig. S12, A and C) without affecting these cells in the VZ (fig. S12B). These results demonstrate that the hTKTL1-induced increase in bRG abundance requires the conversion of acetyl-CoA in malonyl-CoA.

To further analyze the requirement of fatty acid synthesis in the hTKTL1-induced increase

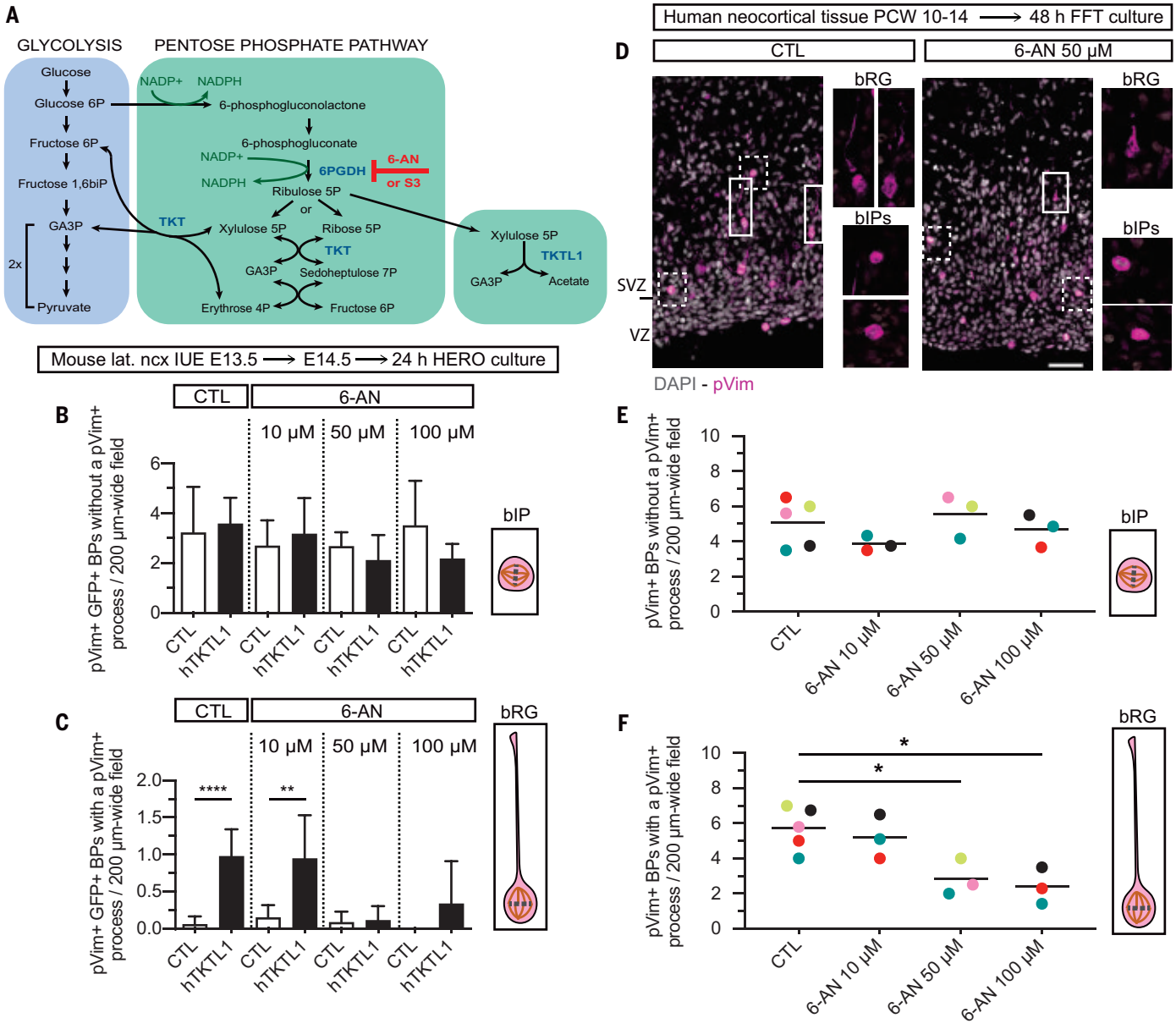


Fig. 5. Inhibition of the pentose phosphate pathway abolishes the hTKTL1-induced increase in bRG in embryonic mouse neocortex and reduces bRG abundance in fetal human neocortex. (**A**) Schematic representation of the glycolysis and pentose phosphate pathways. Known TKT, and proposed TKTL1 (*19*), sites of action are indicated. 6-Aminonicotinamide (6-AN) and S3: 6-phosphogluconate dehydrogenase (6PGDH) inhibitors. (**B** and **C**) Mouse lateral neocortex (lat. ncx.) E13.5 IUE with GFP plasmid, together with empty (control, CTL) or hTKTL1 plasmid. E14.5: Hemisphere rotation culture (HERO, 24 hours) in the absence or presence of 10, 50, or 100 µM 6-AN. Quantifications of pVim$^+$/GFP$^+$/BPs without pVim$^+$ process (mitotic bIPs) (B) and with pVim$^+$ process (mitotic bRG) (C) in 200-µm-wide fields. Means of 3 to 9 embryos. Error bars, SD. (B) Two-way ANOVA; (C) two-way ANOVA with Bonferroni post hoc test, ****$P < 0.0001$, **$P < 0.01$. (**D** to **F**) Free-floating tissue culture (FFT, 48 hours) of human neocortical tissue (PCW 10 to 14) for 48 hours, without or with 10, 50, or 100 µM 6-AN. (D) pVim (magenta) immunofluorescence plus DAPI staining (gray). Right sides: white boxed areas at higher magnification; pVim$^+$/BP without pVim$^+$ process (mitotic bIP, dashed lines), and with pVim$^+$ process (mitotic bRG, solid lines). Scale bar, 40 µm. [(E) and (F)] Quantification of pVim$^+$/BPs without pVim$^+$ process (mitotic bIPs) (E) and with pVim$^+$ process (mitotic bRG) (F) in 200-µm-wide field. Means of 3 to 5 fetal samples, each a different color. (E) One-way ANOVA; (F) one-way ANOVA with Tukey post hoc test, *$P < 0.05$.

in bRG abundance, we explored whether the following step of fatty acid synthesis, the condensation of acetyl-CoA and malonyl-CoA by fatty acid synthase (FAS), is required. Using our standard approach of hTKTL1 IUE at E13.5, E14.5 neocortex was subjected to HERO culture for 24 hours in the absence or presence of the FAS inhibitor Orlistat. Orlistat at 10 and 20 µM completely suppressed the hTKTL1-induced increase in mitotic bRG (fig. S13, A and D) without affecting mitotic bIPs (fig. S13C) and mitotic aRG (fig. S13B). Accordingly, treatment with 10 µM or 20 µM Orlistat completely suppressed the hTKTL1-induced increase in the percentage of GFP$^+$ cells in the SVZ that were Sox2$^+$ (fig. S13, E and G) without affecting these cells in the VZ (fig. S13F). These data therefore demonstrate that hTKTL1 induced the increase in bRG abundance through a fatty acid synthesis–dependent mechanism.

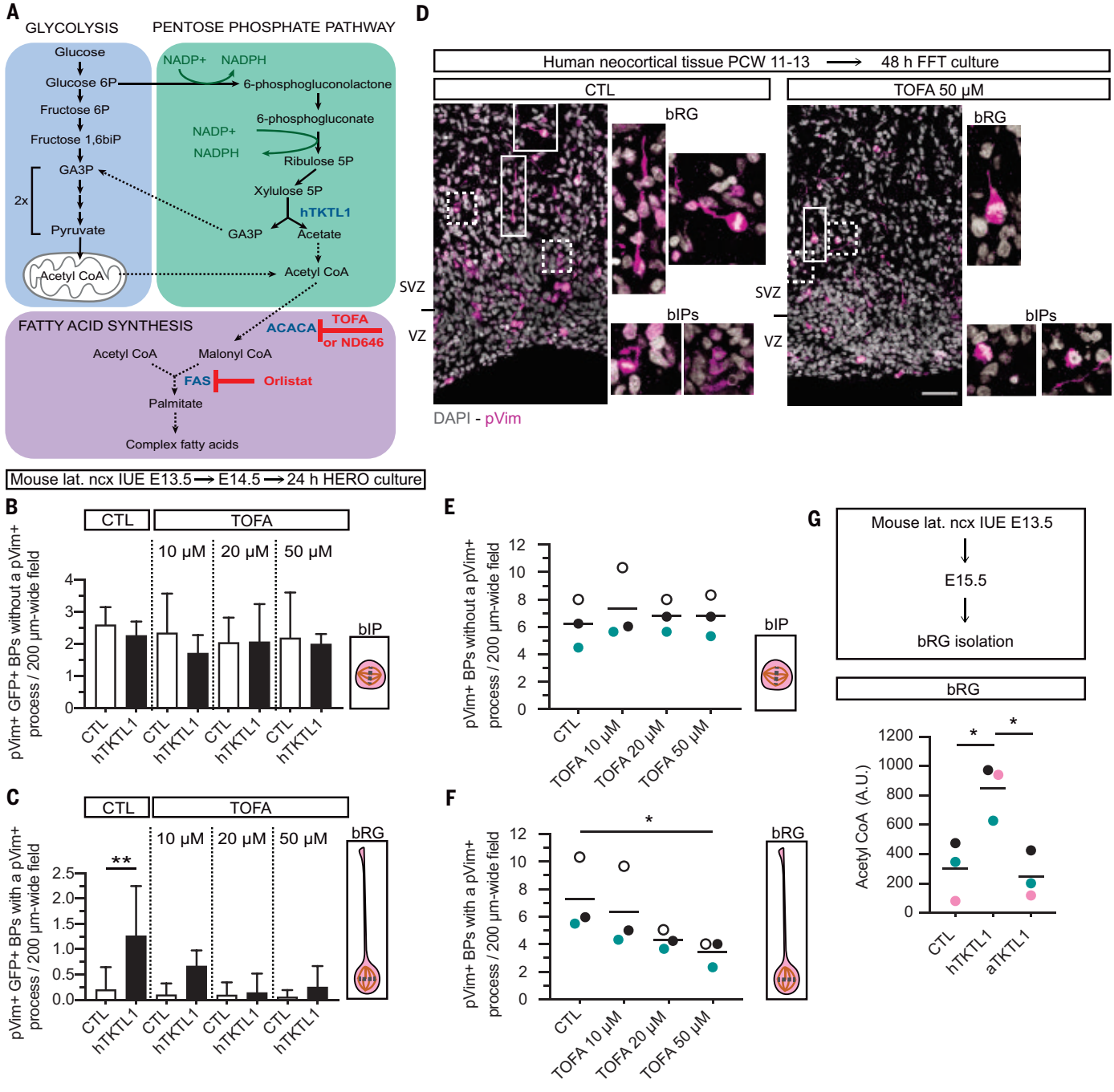


Fig. 6. Inhibition of fatty acid synthesis abolishes the hTKTL1-induced increase in bRG in embryonic mouse neocortex and reduces bRG abundance in fetal human neocortex. (**A**) Schematic representation of the glycolysis, pentose phosphate, and fatty acid synthesis pathways. Proposed TKTL1 site of action (*19*) indicated. 5-(Tetradecyloxy)-2-furoic acid (TOFA) and ND646: acetyl-CoA carboxylase (ACACA, conversion of acetyl-CoA into malonyl-CoA, fatty acid synthesis first and rate-limiting step) inhibitors. Orlistat: fatty acid synthase (FAS) inhibitor. (**B** and **C**) Mouse lateral neocortex E13.5 IUE with GFP plasmid, together with either an empty (control, CTL) or an hTKTL1 plasmid. E14.5: Hemisphere rotation culture (HERO, 24 hours), in the absence or presence of 10, 20, or 50 µM TOFA. Quantification of $pVim^+/GFP^+/BPs$ without $pVim^+$ process (mitotic bIPs) (B) and with $pVim^+$ process (mitotic bRG) (C) in 200-µm-wide field. Means of 5 to 7 embryos. Error bars, SD. (B) Two-way ANOVA; (C) two-way ANOVA with Bonferroni post hoc test, $^{**}P < 0.01$. (**D** to **F**) Free-floating tissue culture (FFT, 48 hours) of human neocortical tissue (PCW 11 to 13), without or with 10, 20, or 50 µM TOFA. (D) pVim (magenta) immunofluorescence plus DAPI staining (gray). Right side: white boxed areas at higher magnification; $pVim^+$/BP without $pVim^+$ process (mitotic bIP, dashed lines), and with $pVim^+$ process (mitotic bRG, solid lines). Scale bar, 50 µm. [(E) and (F)] Quantification of $pVim^+$/BPs without $pVim^+$ process (mitotic bIPs) (E) and with $pVim^+$ process (mitotic bRG) (F) in 200-µm-wide field. Means of 3 fetal samples, each a different color. (E) One-way ANOVA; (F) Friedman test with Dunn post hoc test, $^{*}P < 0.05$. (**G**) Mouse lateral neocortex E13.5 IUE with GFP plasmid, together with either empty (control, CTL), hTKTL1, or aTKTL1 plasmid. E15.5 GFP^+/bRG isolation by FACS. Acetyl-CoA concentration in arbitrary units (A.U., see methods). Means of 3 bRG isolations, each a different color. One-way ANOVA with Tukey post hoc test, $^{*}P < 0.05$.

We also used our mouse test system (neocortex hTKTL1 IUE → HERO culture) to explore potentially additive effects when inhibitors of both the PPP and fatty acid synthesis are used together. 6-AN at 10 μM [which, in light of its affinity for 6PGDH, is likely to partially inhibit the PPP (*34*)] plus 10 μM TOFA partially reduced the hTKTL1-induced increase in the level of bRG (fig. S11, G and J) to a similar extent as 10 μM TOFA alone (Fig. 6C). 6-AN at 50 μM plus 20 μM TOFA completely abolished the hTKTL1-induced increase in the level of bRG (fig. S11J) similarly to 6-AN (Fig. 5C) and TOFA (Fig. 6C) alone. Neither treatment with both inhibitors together affected mitotic bIPs (fig. S11I) and mitotic aRG (fig. S11H). On the likely assumption that partial inhibition of the PPP still provides sufficient acetyl-CoA levels for fatty acid synthesis to occur, the lack of an additive effect of 10 μM 6-AN plus 10 μM TOFA on bRG reduction is consistent with the model (Fig. 6A) that fatty acid synthesis is rate-limiting for the generation of bRG and that the pentose phosphate pathway is upstream of fatty acid synthesis by providing acetyl-CoA.

Fatty acid synthesis inhibition reduces human neocortical bRG

To examine the physiological relevance of these data, we asked whether inhibition of fatty acid synthesis in fetal human neocortex, which endogenously expresses *hTKTL1*, would reduce bRG abundance. To this end, we treated fetal human neocortical tissue in FFT culture for 48 hours with 10, 20, or 50 μM TOFA. Treatment with 50 μM TOFA decreased the abundance of mitotic bRG, identified as process-bearing $pVim^+$ BPs (Fig. 6, D and F), but not that of mitotic bIPs, identified as process-lacking $pVim^+$ BPs (Fig. 6, D and E). Furthermore, human neocortical tissue treated for 48 hours with either (i) 5 μM ND-646 or (ii) 10 or 20 μM Orlistat reduced the abundance of mitotic bRG (fig. S12, D and F, and fig. S13, H and J), but not of mitotic bIPs (fig. S12, D and E, and fig. S13, H and I). The reductions in bRG observed upon TOFA, ND646, and Orlistat treatments were not due to an increase in apoptosis (fig. S11, K and L; fig. S12, G and H; and fig. S13, K and L). Hence, fatty acid synthesis, which is required for the hTKTL1-induced increase in bRG abundance in embryonic mouse neocortex (Fig. 6C and fig. S13D), is also required to maintain the abundance of bRG in fetal human neocortex.

hTKTL1 increases acetyl-CoA concentration in bRG

Our data with the various inhibitors of the PPP and fatty acid synthesis suggest a mechanism in which the increase in bRG abundance by hTKTL1 involves an increase in fatty acid synthesis, which is enabled by hTKTL1's action in the PPP that increases the relevant precursor metabolites for fatty acid synthesis. If so, one would expect an increase in the concentration of acetyl-CoA in bRG upon hTKTL1 expression. We therefore used mass spectrometry to determine the acetyl-CoA concentration in bRG isolated from E15.5 control-, hTKTL1- and aTKTL1-electroporated mouse neocortex (IUE at E13.5). This showed that, indeed, the acetyl-CoA concentration was higher in bRG isolated from hTKTL1-electroporated neocortex than in bRG isolated from control-electroporated neocortex (Fig. 6G). The ability of TKTL1 to increase acetyl-CoA concentration in bRG is restricted to hTKTL1, as bRG isolated from aTKTL1-electroporated neocortex had a similar concentration in acetyl-CoA as the control (Fig. 6G).

Discussion

Our results provide insight into the development of the neocortex of modern humans compared to that of archaic humans. This insight is based on the present findings pertaining to *TKTL1*, a gene belonging to the transketolase family that arose in mammals and that, reflecting its involvement in various types of human cancer, has been implicated in cell proliferation.

hTKTL1 affects bRG, not other neuroprogenitors

The human-specific variant *hTKTL1* increases the abundance, among the various types of neuroprogenitors, of bRG. Upon overexpression in embryonic mouse neocortex, hTKTL1 was found in all types of neuroprogenitors. The selective action of hTKTL1, but not aTKTL1, on bRG abundance has physiological relevance, as KO of *hTKTL1* alone was sufficient to reduce the normal bRG level in fetal human neocortical tissue. The selectivity of hTKTL1 for bRG is likely related to the need of bRG for de novo membrane biogenesis in order to proliferate.

Single amino acid substitution underlies hTKTL1 effect

Previous studies reported on the role of single modern human-specific amino acid substitutions in adenylosuccinate lyase, an enzyme involved in purine metabolism (*35*), and in NOVA1, a splicing regulator involved in neuronal maturation (*36*) [but see (*37*)]. Our results identify the functional role of a single modern human-specific amino acid substitution in the control of the abundance of one type of neuroprogenitor during neocortex development. The lysine-to-arginine substitution that occurred during human evolution (*27*) allows hTKTL1 to ensure the elevated bRG numbers characteristic of fetal human neocortex. Not only the reduction in bRG numbers upon *hTKTL1* KO in fetal human neocortical tissue, but also the lower bRG numbers in cerebral organoids expressing archaic TKTL1, support this conclusion. Our data therefore imply a difference in cortical neuroprogenitor composition between Neanderthals and modern humans.

hTKTL1 action requires the PPP

Our demonstration that the ability of hTKTL1 to ensure elevated bRG numbers requires the PPP adds another case to the emerging concept of a role of metabolism in neuroprogenitor proliferation (*38*) and of human-specific molecular changes affecting this metabolism. Another example is the human-specific gene *ARHGAP11B*, the protein of which acts in mitochondria and promotes BP proliferation by stimulating glutaminolysis (*39*). This ability of ARHGAP11B is based on a single C → G nucleotide substitution (*40*). The ability of hTKTL1 to ensure elevated bRG numbers via its action in the PPP is also based on a single A → G nucleotide substitution that converts the lysine in archaic TKTL1 into the arginine of modern human TKTL1.

Our data show that bRG numbers, in contrast to bIP numbers, depend on the PPP. Given the inability of aTKTL1 to increase bRG abundance, it is worth considering how the single amino acid substitution (lysine to arginine) in TKTL1 during human evolution may have affected its metabolic function. The side chain of arginine is a much stronger base than that of lysine, which may be advantageous for the enzymatic reaction reported for hTKTL1, that is, to cleave xylulose 5-phosphate into glyceraldehyde 3-phosphate and acetate (*19*).

hTKTL1 action requires fatty acid synthesis

The ability of hTKTL1 to ensure elevated bRG numbers via fatty acid synthesis is in agreement with a previous finding that knockdown of *hTKTL1* in the THP-1 leukemia cell line, which expresses high levels of *hTKTL1* mRNA, led to the reduction of the amount of several lipids containing fatty acids (*19*), and with the notion that fatty acid synthesis is a feature of highly proliferative cells, such as cancer cells (*41*). In this context, the role of the essential precursor to fatty acid synthesis, acetyl-CoA, in hTKTL1 action is further supported by our demonstration that hTKTL1, but not aTKTL1, increases acetyl-CoA levels in bRG.

To understand how fatty acid synthesis might be related to bRG proliferation, one should consider that fatty acids are building blocks for membrane synthesis and that the number of bRG cell processes has been shown to be linked to bRG proliferative capacity (*31*). bRG with an increased number of cell processes have been proposed to be able to sense additional extrinsic signals, which in turn activate pro-proliferative signaling pathways. Hence, hTKTL1 is likely to stimulate bRG proliferation, and hence bRG abundance, by promoting cell process growth via synthesis of fatty

acids that are then used to build membranes. In support of this notion, hTKTL1 expression in developing ferret neocortex increased the proportion of bRG with multiple radial processes.

hTKTL1 may influence neocortex shape

Our findings that hTKTL1 expression in developing ferret neocortex increases its lateral extension and gyrus size raise the possibility that hTKTL1 played a role in determining the shape of the neocortex of modern humans. Endocast analyses indicate that Neanderthals and modern humans had/have a similar endocranial volume, consistent with similar brain size (*5*), but that the shape of their brains was/is different. The skull of modern humans has a globular shape, whereas the skull of Neanderthals had a more elongated shape, which is typical of the living apes (*5*). In this context, hTKTL1 could be involved in the formation of a more globular neocortex in modern humans. In addition, modern humans have a larger parietotemporal lobe than Neanderthals (*42*). In support of a possible role of hTKTL1 in neocortical regionalization, it has been reported that *hTKTL1* mRNA expression in PCW 15–21 fetal human neocortex is enriched in the oSVZ of the frontal, as compared to caudal, neocortex (*28*). Consistent with these data, we found an increase in *hTKTL1* mRNA levels in the human frontal lobe compared to the occipital lobe. Hence, hTKTL1 could have played a role in the expansion of the frontal lobe of the neocortex of modern humans.

Neurogenesis in modern humans versus Neanderthals

An implication of the ability of hTKTL1 to ensure elevated bRG numbers pertains to a difference in cortical neuron production between modern humans and Neanderthals. As is evident from our mathematical modeling (Fig. 3, A to C), cortical neuron production over time is greater with bRG than bIPs. In line with this, we observed an increase specifically in late-born (i.e., upper-layer) neurons upon hTKTL1 expression. As such an increase is a hallmark of neocortex development during human evolution (*6*), the selective increase in bRG numbers upon hTKTL1, but not aTKTL1, expression implies that during human evolution, modern humans acquired a more efficient mode of cortical neuron production than Neanderthals. Consistent with this notion, human cerebral organoids expressing archaic TKTL1 showed not only a reduction in bRG, but also in neurons.

The higher *TKTL1* mRNA expression in the frontal lobe than in other parts of the fetal human neocortex suggests that the increase in cortical neuron production pertains, in particular, to the frontal lobe of the neocortex of modern humans. Furthermore, the increase in cortical neuron production, resulting from the elevated bRG numbers due to hTKTL1 action, presumably contributes to the fact that modern humans generate the largest number of cortical neurons of all living primates (*43*).

Conclusions

Our study demonstrates that the gene *hTKTL1* of modern humans is sufficient to increase the abundance of bRG, a neuroprogenitor type that has a role in neocortical evolution. This increase in bRG abundance is induced by the modern human variant of TKTL1 and not by its archaic variant. This suggests that a single human-specific amino acid substitution in TKTL1 underlies changes in cell metabolism that ultimately result in a specific composition of neuroprogenitors and features of cortical neurogenesis that distinguish modern humans from Neanderthals and other extinct archaic humans.

Methods summary

IUE of embryonic mouse neocortex

IUE was performed as described (*31*). E13.5 (for lateral neocortex) or E15.5 (for medial neocortex) embryos were injected intraventricularly with a solution containing 0.1% Fast Green (Sigma) in PBS, 1 μg/μl of pCAGGS plasmid (either empty vector, hTKTL1, or aTKTL1) and 0.4 μg/μl or 1 μg/μl (for cell sorting) of pCAGGS-GFP. The embryos were then electroporated and sacrificed at E15.5 or E17.5 (for lateral neocortex) or at E18.5 (for medial neocortex).

IUE of embryonic ferret neocortex

E33 ferret neocortex was electroporated as described (*44*). The embryos were injected intraventricularly with 0.1% Fast Green (Sigma) in PBS, 2.5 μg/μl of either empty pCAGGS (control) or pCAGGS-hTKTL1, and 1 μg/μl of either pCAGGS-GFP or pCAGGS-mCherry. The injected embryos were electroporated. After birth, the pups were sacrificed at P2 or P16.

Mouse cerebral hemisphere rotation culture

Cerebral hemispheres were dissected from E14.5 mouse embryos 24 hours after IUE at E13.5, and the electroporated hemispheres were placed into hemisphere rotation (HERO) culture as described (*45*), with minor modifications. Hemispheres were cultured with mouse slice culture medium containing 0.1% DMSO and either no addition (control) or one of the inhibitors, for 24 hours at 37°C, in a whole-embryo culture incubator (Ikemoto Scientific Technology).

Human neocortex free-floating tissue culture

Free-floating tissue (FFT) culture of fetal human neocortical tissue (PCW 8–14) was performed for 48 hours (inhibitors) or 72 hours (CRISPR/Cas9 KO) as described (*46*). For the pharmacological inhibitor studies, the tissue was incubated with 0.1% DMSO and either no further addition (control) or one of the inhibitors. The flasks were incubated for 48 or 72 hours at 37°C in a whole-embryo culture incubator (Ikemoto Scientific Technology).

Human fetal neocortical tissue electroporation

Ex vivo electroporation of human fetal neocortical tissue was performed as described (*39*). Human fetal neocortical tissue was electroporated using a mixture of either the Cas9–LacZ-gRNA complex (control) or the Cas9–hTKTL1-gRNA complex, Fast Green, pCAGGS-GFP, and glycerol, followed by 72-hour FFT culture.

Cerebral organoids

Two mock-edited (hTKTL1) and two gene-edited (aTKTL1) H9 cell lines were differentiated into cerebral organoids using previously published protocols (*47*, *48*). The cerebral organoids were fixed at day 50 with PFA at 4°C for 2 hours.

Cryosectioning and immunofluorescence

Cryosectioning of fixed tissues (including cerebral organoids) was performed as described (*31*). Immunofluorescence of cryosections subjected to antigen retrieval was performed as described (*29*).

bRG isolation and determination of acetyl-CoA

We used a modification of a published procedure (*22*) to isolate bRG from E15.5 mouse neocortices, identified as GLAST-positive and prominin-1–negative cells, after IUE at E13.5 (GFP plus either empty vector, hTKTL1, or aTKTL1). bRG were isolated from the neocortical cell suspension by FACS. Acetyl-CoA levels in the FACS-isolated bRG were determined by mass spectrometry.

Other methods

All other methods used were carried out according to standard procedures.

REFERENCES AND NOTES

1. M. Breyl, Triangulating Neanderthal cognition: A tale of not seeing the forest for the trees. *Wiley Interdiscip. Rev. Cogn. Sci.* **12**, e1545 (2021). doi: 10.1002/wcs.1545; pmid: 32918796
2. D. L. Hoffmann *et al*., U-Th dating of carbonate crusts reveals Neandertal origin of Iberian cave art. *Science* **359**, 912–915 (2018). doi: 10.1126/science.aap7778; pmid: 29472483
3. D. Leder *et al*., A 51,000-year-old engraved bone reveals Neanderthals' capacity for symbolic behaviour. *Nat. Ecol. Evol.* **5**, 1273–1282 (2021). doi: 10.1038/s41559-021-01487-z; pmid: 34226702
4. P. Rakic, Evolution of the neocortex: A perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009). doi: 10.1038/nrn2719; pmid: 19763105
5. P. Gunz *et al*., Neandertal introgression sheds light on modern human endocranial globularity. *Curr. Biol.* **29**, 120–127.e5 (2019). doi: 10.1016/j.cub.2018.10.065; pmid: 30554901
6. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011). doi: 10.1016/j.cell.2011.06.030; pmid: 21729779
7. E. Taverna, M. Götz, W. B. Huttner, The cell biology of neurogenesis: Toward an understanding of the development and evolution of the neocortex. *Annu. Rev. Cell Dev. Biol.* **30**,

465–502 (2014). doi: 10.1146/annurev-cellbio-101011-155801; pmid: 25000993
8. I. Kelava *et al*., Abundant occurrence of basal radial glia in the subventricular zone of embryonic neocortex of a lissencephalic primate, the common marmoset Callithrix jacchus. *Cereb. Cortex* **22**, 469–481 (2012). doi: 10.1093/cercor/bhr301; pmid: 22114084
9. X. Wang, J. W. Tsai, B. LaMonica, A. R. Kriegstein, A new subtype of progenitor cell in the mouse embryonic neocortex. *Nat. Neurosci.* **14**, 555–561 (2011). doi: 10.1038/nn.2807; pmid: 21478886
10. A. Shitamukai, D. Konno, F. Matsuzaki, Oblique radial glial divisions in the developing mouse neocortex induce self-renewing progenitors outside the germinal zone that resemble primate outer subventricular zone progenitors. *J. Neurosci.* **31**, 3683–3695 (2011). doi: 10.1523/JNEUROSCI.4773-10.2011; pmid: 21389223
11. M. Betizeau *et al*., Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013). doi: 10.1016/j.neuron.2013.09.032; pmid: 24139044
12. S. A. Fietz *et al*., OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat. Neurosci.* **13**, 690–699 (2010). doi: 10.1038/nn.2553; pmid: 20436478
13. D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010). doi: 10.1038/nature08845; pmid: 20154730
14. V. Borrell, M. Götz, Role of radial glial cells in cerebral cortex folding. *Curr. Opin. Neurobiol.* **27**, 39–46 (2014). doi: 10.1016/j.conb.2014.02.007; pmid: 24632307
15. A. Pinson, W. B. Huttner, Neocortex expansion in development and evolution—from genes to progenitor cell biology. *Curr. Opin. Cell Biol.* **73**, 9–18 (2021). doi: 10.1016/j.ceb.2021.04.008; pmid: 34098196
16. J. F. Coy *et al*., Molecular cloning of tissue-specific transcripts of a transketolase-related gene: Implications for the evolution of new vertebrate genes. *Genomics* **32**, 309–316 (1996). doi: 10.1006/geno.1996.0124; pmid: 8838793
17. B. L. Horecker, The pentose phosphate pathway. *J. Biol. Chem.* **277**, 47965–47971 (2002). doi: 10.1074/jbc.X200007200; pmid: 12403765
18. J. F. Coy, D. Dressler, J. Wilde, P. Schubert, Mutations in the transketolase-like gene TKTL1: Clinical implications for neurodegenerative diseases, diabetes and cancer. *Clin. Lab.* **51**, 257–273 (2005). pmid: 15991799
19. S. Diaz-Moralli *et al*., A key role for transketolase-like 1 in tumor metabolic reprogramming. *Oncotarget* **7**, 51875–51897 (2016). doi: 10.18632/oncotarget.10429; pmid: 27391434
20. A. A. Pollen *et al*., Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015). doi: 10.1016/j.cell.2015.09.004; pmid: 26406371
21. S. A. Fietz *et al*., Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11836–11841 (2012). doi: 10.1073/pnas.1209647109; pmid: 22753484
22. M. Florio *et al*., Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465–1470 (2015). doi: 10.1126/science.aaa1975; pmid: 25721503
23. M. B. Johnson *et al*., Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat. Neurosci.* **18**, 637–646 (2015). doi: 10.1038/nn.3980; pmid: 25734491
24. W. Sun *et al*., TKTL1 is activated by promoter hypomethylation and contributes to head and neck squamous cell carcinoma carcinogenesis through increased aerobic glycolysis and HIF1α stabilization. *Clin. Cancer Res.* **16**, 857–866 (2010). doi: 10.1158/1078-0432.CCR-09-2604; pmid: 20103683
25. R. Peltonen *et al*., High TKTL1 expression as a sign of poor prognosis in colorectal cancer with synchronous rather than metachronous liver metastases. *Cancer Biol. Ther.* **21**, 826–831 (2020). doi: 10.1080/15384047.2020.1803008; pmid: 32795237
26. W. Yuan *et al*., Silencing of TKTL1 by siRNA inhibits proliferation of human gastric cancer cells in vitro and in vivo. *Cancer Biol. Ther.* **9**, 710–716 (2010). doi: 10.4161/cbt.9.9.11431; pmid: 20200485
27. K. Prüfer *et al*., The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi: 10.1038/nature12886; pmid: 24352235
28. J. A. Miller *et al*., Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
29. S. Vaid *et al*., A novel population of Hopx-dependent basal radial glial cells in the developing mouse neocortex. *Development* **145**, dev169276 (2018). doi: 10.1242/dev.169276; pmid: 30266827
30. M. A. Martínez-Martínez *et al*., A restricted period for formation of outer subventricular zone defined by Cdh1 and Trnp1 levels. *Nat. Commun.* **7**, 11812 (2016). doi: 10.1038/ncomms11812; pmid: 27264089
31. N. Kalebic *et al*., Neocortical expansion due to increased proliferation of basal progenitors is linked to changes in their morphology. *Cell Stem Cell* **24**, 535–550.e9 (2019). doi: 10.1016/j.stem.2019.02.017; pmid: 30905618
32. N. Matsumoto, Y. Shinmyo, Y. Ichikawa, H. Kawasaki, Gyrification of the cerebral cortex requires FGF signaling in the mammalian brain. *eLife* **6**, e29285 (2017). doi: 10.7554/eLife.29285; pmid: 29132503
33. V. Fernández, C. Llinares-Benadero, V. Borrell, Cerebral cortex expansion and folding: What have we learned? *EMBO J.* **35**, 1021–1044 (2016). doi: 10.15252/embj.201593701; pmid: 27056680
34. J. S. Hothersall, M. Gordge, A. A. Noronha-Dutra, Inhibition of NADPH supply by 6-aminonicotinamide: Effect on glutathione, nitric oxide and superoxide in J774 cells. *FEBS Lett.* **434**, 97–100 (1998). doi: 10.1016/S0014-5793(98)00959-4; pmid: 9738459
35. V. Stepanova *et al*., Reduced purine biosynthesis in humans after their divergence from Neandertals. *eLife* **10**, e58741 (2021). doi: 10.7554/eLife.58741; pmid: 33942714
36. C. A. Trujillo *et al*., Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment. *Science* **371**, eaax2537 (2021). doi: 10.1126/science.aax2537; pmid: 33574182
37. T. Maricic *et al*., Comment on "Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment". *Science* **374**, eabi6060 (2021). doi: 10.1126/science.abi6060; pmid: 34648345
38. T. Namba, J. Nardelli, P. Gressens, W. B. Huttner, Metabolic regulation of neocortical expansion in development and evolution. *Neuron* **109**, 408–419 (2021). doi: 10.1016/j.neuron.2020.11.014; pmid: 33306962
39. T. Namba *et al*., Human-specific ARHGAP11B acts in mitochondria to expand neocortical progenitors by glutaminolysis. *Neuron* **105**, 867–881.e9 (2020). doi: 10.1016/j.neuron.2019.11.027; pmid: 31883789
40. M. Florio, T. Namba, S. Pääbo, M. Hiller, W. B. Huttner, A single splice site mutation in human-specific *ARHGAP11B* causes basal progenitor amplification. *Sci. Adv.* **2**, e1601941 (2016). doi: 10.1126/sciadv.1601941; pmid: 27957544
41. F. Röhrig, A. Schulze, The multifaceted roles of fatty acid synthesis in cancer. *Nat. Rev. Cancer* **16**, 732–749 (2016). doi: 10.1038/nrc.2016.89; pmid: 27658529
42. M. Bastir, A. Rosas, D. E. Lieberman, P. O'Higgins, Middle cranial fossa anatomy and the origin of modern humans. *Anat. Rec.* **291**, 130–140 (2008). doi: 10.1002/ar.20636; pmid: 18213701
43. S. Herculano-Houzel, The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl. 1), 10661–10668 (2012). doi: 10.1073/pnas.1201895109; pmid: 22723358
44. N. Kalebic *et al*., Human-specific *ARHGAP11B* induces hallmarks of neocortical expansion in developing ferret neocortex. *eLife* **7**, e41241 (2018). doi: 10.7554/eLife.41241; pmid: 30484771
45. J. Schenk, M. Wilsch-Bräuninger, F. Calegari, W. B. Huttner, Myosin II is required for interkinetic nuclear migration of neural progenitors. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16487–16492 (2009). doi: 10.1073/pnas.0908928106; pmid: 19805325
46. K. R. Long *et al*., Extracellular matrix components HAPLN1, lumican, and collagen I cause hyaluronic acid-dependent folding of the developing human neocortex. *Neuron* **99**, 702–719.e6 (2018). doi: 10.1016/j.neuron.2018.07.013; pmid: 30078576
47. M. A. Lancaster, J. A. Knoblich, Generation of cerebral organoids from human pluripotent stem cells. *Nat. Protoc.* **9**, 2329–2340 (2014). doi: 10.1038/nprot.2014.158; pmid: 25188634
48. F. Mora-Bermúdez *et al*., Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife* **5**, e18683 (2016). doi: 10.7554/eLife.18683; pmid: 27669147

ACKNOWLEDGMENTS

We apologize to all researchers whose work could not be cited because of space limitations. We thank I. Reichardt and M. Sarov for their help with the CRISPR-Cas9–mediated *hTKTL1* KO, notably the design and testing of the gRNAs; K. Hackmann and E. Schröck for karyotyping of H9 cell lines; P. Kanis, D. Macak, C. Heide, and H. Zeberg for help and discussions; the members of the Huttner lab for useful discussions; C. Haffner for technical assistance; the services and facilities of the Max Planck Institute of Molecular Cell Biology and Genetics for their technical support, notably J. Peychl and his team of the Light Microscopy Facility, J. Helppi and his team of the Biomedical Services (BMS), and J. Jarrells, I. Nuesslein, and their team of the Cell Technology facility; and A. Münch-Wuttke for the ferret intracardiac perfusion. **Funding:** Supported by the NOMIS Foundation (S.P.), a grant from ERA-NET NEURON (MicroKin) (W.B.H.), and the Max Planck Society (S.P and W.B.H.). **Author contributions:** Conceptualization: A.P., W.B.H.; formal analysis: A.P., L.X.; C.E.O., S.T., T.M.; investigation: A.P., L.X., T.N., N.K., J.P., C.E.O., S.T., K.R., S.R., T.M.; resources: S.R., S.P., R.D., P.W.; writing–original draft: A.P., W.B.H.; writing–review and editing: A.P., W.B.H. with comments from N.K, T.N., L.X., S.P.; project administration and supervision: W.B.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** No dataset or code was generated in this study. The previously published datasets used in this study are publicly available at the NCBI Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo) with accession numbers GSE38805, GSE65000, and GSE66217. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl6422
Materials and Methods
Figs. S1 to S13
References (*49*–*57*)

Submitted 27 July 2021; resubmitted 9 June 2022
Accepted 28 July 2022
10.1126/science.abl6422

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Science Advances
AAAS

GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

RESEARCH ARTICLE SUMMARY

CLIMATE CHANGE

Exceeding 1.5°C global warming could trigger multiple climate tipping points

David I. Armstrong McKay*, Arie Staal, Jesse F. Abrams, Ricarda Winkelmann, Boris Sakschewski, Sina Loriani, Ingo Fetzer, Sarah E. Cornell, Johan Rockström, Timothy M. Lenton*

INTRODUCTION: Climate tipping points (CTPs) are a source of growing scientific, policy, and public concern. They occur when change in large parts of the climate system—known as tipping elements—become self-perpetuating beyond a warming threshold. Triggering CTPs leads to significant, policy-relevant impacts, including substantial sea level rise from collapsing ice sheets, dieback of biodiverse biomes such as the Amazon rainforest or warm-water corals, and carbon release from thawing permafrost. Nine policy-relevant tipping elements and their CTPs were originally identified by Lenton *et al.* (2008). We carry out the first comprehensive reassessment of all suggested tipping elements, their CTPs, and the timescales and impacts of tipping. We also highlight steps to further improve understanding of CTPs, including an expert elicitation, a model intercomparison project, and early warning systems leveraging deep learning and remotely sensed data.

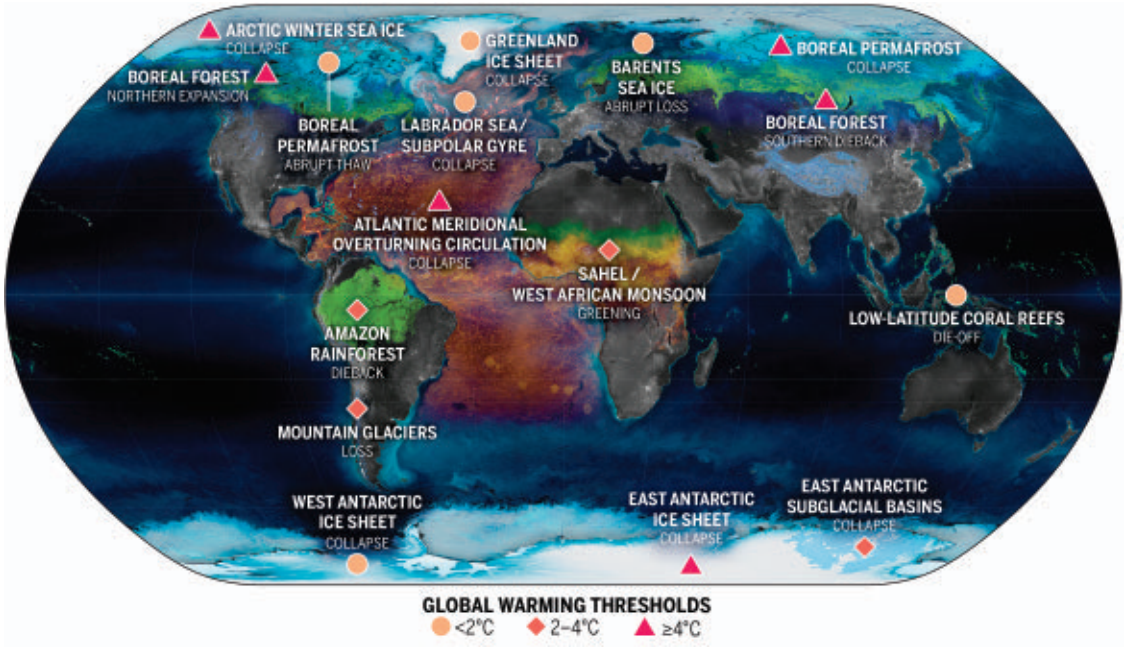
RATIONALE: Since the original identification of tipping elements there have been substantial advances in scientific understanding from paleoclimate, observational, and model-based studies. Additional tipping elements have been proposed (e.g., parts of the East Antarctic ice sheet) and the status of others (e.g., Arctic summer sea ice) has been questioned. Observations have revealed that parts of the West Antarctic ice sheet may have already passed a tipping point. Potential early warning signals of the Greenland ice sheet, Atlantic Meridional Overturning Circulation, and Amazon rainforest destabilization have been detected. Multiple abrupt shifts have been found in climate models. Recent work has suggested that up to 15 tipping elements are now active (Lenton *et al.*, 2019). Hence it is timely to synthesize this new knowledge to provide a revised shortlist of potential tipping elements and their CTP thresholds.

RESULTS: We identify nine global "core" tipping elements which contribute substantially to Earth system functioning and seven regional "impact" tipping elements which contribute substantially to human welfare or have great value as unique features of the Earth system (see figure). Their estimated CTP thresholds have significant implications for climate policy: Current global warming of ~1.1°C above pre-industrial already lies within the lower end of five CTP uncertainty ranges. Six CTPs become likely (with a further four possible) within the Paris Agreement range of 1.5 to <2°C warming, including collapse of the Greenland and West Antarctic ice sheets, die-off of low-latitude coral reefs, and widespread abrupt permafrost thaw. An additional CTP becomes likely and another three possible at the ~2.6°C of warming expected under current policies.

CONCLUSION: Our assessment provides strong scientific evidence for urgent action to mitigate climate change. We show that even the Paris Agreement goal of limiting warming to well below 2°C and preferably 1.5°C is not safe as 1.5°C and above risks crossing multiple tipping points. Crossing these CTPs can generate positive feedbacks that increase the likelihood of crossing other CTPs. Currently the world is heading toward ~2 to 3°C of global warming; at best, if all net-zero pledges and nationally determined contributions are implemented it could reach just below 2°C. This would lower tipping point risks somewhat but would still be dangerous as it could trigger multiple climate tipping points. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: d.mckay@exeter.ac.uk (D.I.A.M.); t.m.lenton@exeter.ac.uk (T.M.L.)
Cite this article as D. I. Armstrong McKay *et al.*, *Science* **377**, eabn7950 (2022). DOI: 10.1126/science.abn7950

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abn7950



The location of climate tipping elements in the cryosphere (blue), biosphere (green), and ocean/atmosphere (orange), and global warming levels at which their tipping points will likely be triggered. Pins are colored according to our central global warming threshold estimate being below 2°C, i.e., within the Paris Agreement range (light orange, circles); between 2 and 4°C, i.e., accessible with current policies (orange, diamonds); and 4°C and above (red, triangles).

RESEARCH ARTICLE

CLIMATE CHANGE

Exceeding 1.5°C global warming could trigger multiple climate tipping points

David I. Armstrong McKay[1,2,3,4]*, Arie Staal[1,2,5], Jesse F. Abrams[3], Ricarda Winkelmann[6], Boris Sakschewski[6], Sina Loriani[6], Ingo Fetzer[1,2], Sarah E. Cornell[1,2], Johan Rockström[1,6], Timothy M. Lenton[3]*

Climate tipping points occur when change in a part of the climate system becomes self-perpetuating beyond a warming threshold, leading to substantial Earth system impacts. Synthesizing paleoclimate, observational, and model-based studies, we provide a revised shortlist of global "core" tipping elements and regional "impact" tipping elements and their temperature thresholds. Current global warming of ~1.1°C above preindustrial temperatures already lies within the lower end of some tipping point uncertainty ranges. Several tipping points may be triggered in the Paris Agreement range of 1.5 to <2°C global warming, with many more likely at the 2 to 3°C of warming expected on current policy trajectories. This strengthens the evidence base for urgent action to mitigate climate change and to develop improved tipping point risk assessment, early warning capability, and adaptation strategies.

Climate tipping points (CTPs) have emerged as a growing research topic and source of public concern (*1*–*3*). Tipping points are defined as "a critical threshold at which a tiny perturbation can qualitatively alter the state or development of a system" (*1*). Several large-scale Earth system components, termed tipping elements, were identified with evidence for tipping points that could be triggered by human activities this century. The initial shortlist constituted Arctic summer sea ice, the Greenland ice sheet (GrIS), the West Antarctic ice sheet (WAIS), Atlantic Meridional Overturning Circulation (now AMOC, previously THC), the El Niño Southern Oscillation, the Indian Summer monsoon, the Sahara/Sahel and West African Monsoon, the Amazon rainforest (AMAZ), and boreal forest. A literature review (*1*) and corresponding expert elicitation (*4*) provided early estimates of the temperature thresholds and potential interactions of these tipping elements. Subsequent work showed how recognition of CTPs considerably affects risk analysis and supports measures to minimize global warming to the Paris target of 1.5°C (*5*, *6*).

Since these early estimates (*1*), there have been considerable advances in our knowledge of CTPs including observations of nonlinear changes in the climate system, statistical early warning methods, paleoclimate evidence, upgraded Earth system models (ESMs), and improved offline models of particular elements (e.g., ice sheets and vegetation). Notably, observations and models suggest that parts of the WAIS may be approaching (*7*, *8*) or have even passed a tipping point (*9*, *10*). Early warning indicators have revealed potential destabilization of the GrIS, AMOC, and AMAZ (*11*–*13*). However, many ESMs still lack processes important for resolving potential tipping behavior—e.g., bias toward AMOC stability (*14*)—or underestimating current tropical carbon sink declines (*15*). Potential causal interactions among tipping elements (*4*) are such that overall tipping of one element increases the likelihood of tipping others (*16*), possibly risking a "tipping cascade" of impacts that may further amplify global warming (*2*, *3*). In the worst case scenario, interactions might produce a global CTP (*3*).

The list of tipping elements has evolved over time (*1*–*3*, *5*) (table S1). Different studies have proposed potential additions including southwest North America, the Yedoma permafrost region, the North Atlantic subpolar gyre (*17*), low-latitude coral reefs, the East Antarctic Ice Sheet (EAIS), Arctic winter sea ice (AWSI), Alpine glaciers (*5*), the northern polar jet stream (*3*), the Congo rainforest (*18*), and the Wilkes and Aurora subglacial basins in East Antarctica (*2*). A range of abrupt shifts have been identified in CMIP5 models (*19*), some of which are not in elements on the original shortlist such as boreal tundra or Antarctic sea ice. Conversely, arguments have been made that Arctic summer sea ice (*20*, *21*), El Niño–Southern Oscillation (ENSO) (*22*, *23*), and monsoons (*24*) should not be classified as CTPs. Numerous temperature threshold estimates have been made since (*1*) with some being revised markedly downward—notably WAIS (*2*, *25*). The recent the Intergovernmental Panel on Climate Change (IPCC) AR6 WG1 report identifies up to 15 candidates [table 4.10 in (*23*)] but was not explicit about their temperature thresholds (*23*).

Here we reassess the climate tipping elements based on the substantial literature published since (*1*), focusing on those triggerable by global warming. We clarify the definition of tipping elements and points and propose a new categorization separating global "core" and regional "impact" tipping elements. We then provide an updated list and assessment of the global mean surface temperature (GMST) range at which each candidate CTP could occur as well as their timescales and climate impacts. Finally we combine this information to assess the likelihood of triggering CTPs at successive global warming levels.

Defining tipping points and tipping elements

Given multiple inconsistent definitions of a CTP in the literature, we anchor on the technical definition provided by (*1*): A tipping point is a threshold in a (forcing) "control parameter" at which a small additional perturbation (within natural variability of ~0.2°C) causes a qualitative change [significantly larger than the standard deviation of natural variability in (*1*)] in the future state of a system [see (*1*) and SM for the full definition]. Here, our specific definition is as follows: Tipping points occur when change in part of the climate system becomes (i) self-perpetuating beyond (ii) a warming threshold as a result of asymmetry in the relevant feedbacks, leading to (iii) substantial and widespread Earth system impacts. We now explain key aspects of this definition in more detail.

Self-perpetuating change

Self-perpetuation mechanisms are critical to the existence of a tipping point in a system, beyond which they propel qualitative change such that even if forcing of the system ceases the qualitative change usually continues to unfold regardless (*20*). IPCC AR6 sometimes uses tipping point to refer to a class of abrupt change in which the subsequent rate of change is independent of the forcing [1.4.4.3 of (*26*)], although this is not part of AR6's core CTP definition [4.7.2 of (*23*)]. Self-perpetuation is usually due to positive feedback within a system attaining sufficient strength to overcome stabilizing negative feedbacks and (temporarily) reach a "runaway" condition (in which an initial change propagating around a feedback loop gives rise to an additional change that is at least as large as the initial change and so on). Most positive feedbacks never attain this condition and instead simply amplify the original driver in a constrained way. Notably, Arctic summer sea ice loss involves the positive ice-albedo feedback, but unlike year-round sea ice loss, that feedback alone is not strong enough to produce a clear threshold

[1]Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. [2]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden. [3]Global Systems Institute, University of Exeter, Exeter, UK. [4]Georesilience Analytics, Leatherhead, UK. [5]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands. [6]Potsdam Institute for Climate Impact Research, Potsdam, Germany.
*Corresponding author. Email: d.mckay@exeter.ac.uk (D.I.A.M.); t.m.lenton@exeter.ac.uk (T.M.L.)

beyond which loss would continue even if global warming stopped (*20*, *21*). Consequently, we describe such feedbacks as "threshold-free".

(Ir)reversibility

Tipping points usually lead to irreversible qualitative change but reversible tipping points are possible as a special case (*1*). Many tipping points result from crossing bifurcation points or attraction basin boundaries in bistable systems, with the resulting hysteresis making tipping effectively irreversible on human timescales. However, self-perpetuating change can also occur across noncatastrophic thresholds in unistable systems (*27*) (supplementary text S1). Other definitions of CTPs are more restrictive and require irreversibility, for example: "a system reorganizes… and does not return to the initial state even if the drivers of the change are abated" [6.1.1 of (*22*)]. The IPCC AR6 does not require irreversibility as this is difficult to prove for long timescales given model limitations: "A tipping point is a critical threshold beyond which a system reorganizes, often abruptly and/or irreversibly" [4.7.2 of (*23*)]. AR6 uses abruptness and irreversibility as proxies for tipping dynamics but does not specify criteria for system reorganization and sometimes does not clearly differentiate which abrupt and/or irreversible changes are considered tipping points (e.g., irreversible ocean temperature change is listed alongside potential tipping points in table 4.10 of (*23*) and box 12.1 in table 1 of (*28*) but has no clear critical threshold).

Timescale and abruptness

We allow for CTPs (e.g., in ice sheets) in which the resulting qualitative change is slower than the anthropogenic forcing causing it—i.e., not abrupt in the sense defined as faster than the cause (*29*). We only require that the transition to a new state occurs at a rate determined by the climate (sub)system itself (*29*). The resulting committed (often irreversible) qualitative changes can unfold over centuries to millennia [here we relax the ethical time horizon of (*1*) from ~1 thousand years (ky) to ~10 ky], but crucially they can increase short-term impacts (e.g., rate of sea level rise). Other authors require a tipping point to produce abrupt change (*30*) thereby excluding events such as ice sheet collapse. The IPCC defines abrupt change as "substantially faster than the rate of change in recent history" in AR6 [1.4.4.3 of (*26*)], which could allow for slower changes than anthropogenic forcing. However, AR6 also gave a more restrictive timescale-based definition for abrupt climate change as taking place over a few decades or less (i.e., as fast as anthropogenic forcing) and persisting for at least a few decades [4.7.2 of (*23*)]. More than a dozen abrupt changes have been found in CMIP5 model output (*19*) (table S2) but such changes could simply be the result of an abrupt change in forcing without involving CTPs. Below we assess which abrupt changes indicate potential tipping elements and which do not involve self-perpetuating feedback.

Spatial scale

Tipping elements are defined as components of the Earth system that are at least subcontinental in scale (of the order of 1000 km, i.e., ~1 M km_2) and could pass a tipping point as a result of actions this century (*1*). If self-perpetuating change (and a corresponding tipping point) occurs at a subcontinental scale then this qualifies as a global core tipping element. However, there are many examples of runaway feedback and associated tipping points at smaller spatial scales. Where a change in forcing (e.g., temperature) is fairly uniform across a large spatial scale, such that a smaller-scale tipping point is crossed near-synchronously in many locations that span a subcontinental scale (e.g., coral bleaching across the Great Barrier Reef or committed loss of Himalayan glaciers), then these are considered potential regional impact tipping elements. However, where systems exhibit localized tipping points (1 m to 1 km) at different forcing levels such that change does not self-perpetuate beyond a clear shared threshold (e.g., methane hydrates), these are classed as threshold-free feedbacks because the accumulated global consequences of multiple localized tipping events remain roughly proportional to the forcing.

Impacts

Climate tipping elements in (*1*) either (i) contribute significantly to the overall mode of operation of the Earth system (such that tipping them modifies the overall state of the whole system), (ii) contribute significantly to human welfare (such that tipping them affects >~100 million people), or (iii) have great value in themselves as a unique feature of the Earth system [expanded from the biosphere used in (*1*)]. Global core tipping elements must meet criterion (i) whereas regional impact tipping elements must meet criterion (ii) or (iii) but not (i). Regarding (i), crossing a tipping point need not involve feedback to global atmospheric composition or temperature—self-perpetuating feedback can exist entirely within a tipping element (*1*)—but there is usually causal coupling to other tipping elements such as through heat, salt, water, carbon, or momentum fluxes (*4*). Often there is feedback to global warming and where this exceeds ±0.1°C (i.e., natural variability and the triggering perturbation) we consider this to meet criterion (i). Thus, near-synchronous, large-scale crossing of smaller-scale tipping points can qualify as a global core tipping element if it changes warming by >0.1°C.

The climate tipping elements

Based on current observations, paleorecords, and model runs subsequent to (*1*), we draw up a longlist of proposed climate tipping elements. Together with expert judgment for each proposed element, we summarize the evidence and confidence levels for self-perpetuation, temperature thresholds, hysteresis or irreversibility, transition timescales, and global or regional impacts on climate (Materials and Methods, table S3, and supplementary text S2). Based on this evidence and the definitions in the preceding section, we shortlist global core and regional impact climate tipping elements (Table 1 and Fig. 1). Other candidate tipping elements that we consider uncertain, unlikely, or that have threshold-free feedbacks are discussed in the supplementary text along with differences with past assessments (table S4).

Cryosphere

Arctic sea ice (AWSI/BARI)

An abrupt collapse in AWSI (*31*) is observed in some CMIP5 models beyond ~4.5°C (*19*, *32*), which arises either from asymmetry in ice formation and loss timescales creating a threshold response or from local positive feedback cycles. Hence we class AWSI as a global core tipping element (medium confidence), with a best estimate threshold of ~6.3°C (4.5 to 8.7°C, based on CMIP5) (high confidence), timescales of 20 years (10 to 100 years) (high confidence), and GMST feedback of ~+0.6°C (high confidence) [~+0.25°C when free of summer ice; regional ~+0.6 to 1.2°C (low confidence)]. A subcase is abrupt loss of Barents Sea winter ice (BARI), which occurs at ~1.6°C in two CMIP5 models (*19*), is self-reinforced by an increased inflow of warm Atlantic waters (*33*), and has substantial impacts on atmospheric circulation, European climate, and potentially the AMOC (*34*). We consider BARI a probable regional impact tipping element (medium confidence) with a threshold of 1.6°C (1.5 to 1.7°C) (low confidence), a timescale of ~25 years (low confidence), and regional warming (high confidence). By contrast Arctic summer sea ice (ASSI)—despite declining rapidly since the 1970s and outpacing previous IPCC projections since the 1990s—is responding linearly to cumulative emissions (*35*). This decline is amplified by the ice-albedo feedback and possibly feedbacks to cloud cover but damped by negative heat loss feedbacks (*20*). CMIP6 models better capture historical ASSI decline and project that ice-free Septembers will occur occasionally above 1.5°C GMST, become common beyond 2°C, and remain permanent at ~3°C (*36*). However, the linear modeled and observed responses suggest that ASSI is unlikely to feature a tipping point beyond which loss would self-perpetuate (*21*, *36*). Hence, we recategorize ASSI as a threshold-free feedback.

Greenland ice sheet (GrIS)

The GrIS is shrinking at an accelerated rate as a result of both net surface melt and accelerated

Table 1. Table showing our literature-based threshold, timescale, and impact estimates for the tipping elements we categorize as global core or regional impact. Element acronym colors indicate Earth system domain (blue, cryosphere; green, biosphere; orange, ocean-atmosphere), and element name and estimate colors indicate subjective confidence levels (green, high; yellow, medium; red, low). Bolded element names indicate elements featured in previous climate tipping element characterizations.

Category	Proposed climate tipping element (and tipping point)		Threshold (°C)			Timescale (years)			Maximum impact* (°C)	
			Est.	Min	Max	Est.	Min	Max	Global	Region
Global core tipping elements	GrIS	**Greenland Ice Sheet (collapse)**	1.5	0.8	3.0	10k	1k	15k	0.13	0.5 to 3.0
	WAIS	**West Antarctic Ice Sheet (collapse)**	1.5	1.0	3.0	2k	500	13k	0.05	1.0
	LABC	Labrador-Irminger Seas / SPG Convection (collapse)	1.8	1.1	3.8	10	5	50	-0.50	-3.0
	EASB	East Antarctic Subglacial Basins (collapse)	3.0	2.0	6.0	2k	500	10k	0.05	?
	AMAZ	**Amazon Rainforest (dieback)**	3.5	2.0	6.0	100	50	200	Partial: 30 GtC / 0.1°C Total: 75 GtC / 0.2°C	0.4 to 2.0
	PFTP	**Boreal Permafrost (collapse)**	4.0	3.0	6.0	50	10	300	125-250 GtC / 175-350 GtCe / 0.2-0.4°C	~
	AMOC	**Atlantic M.O. Circulation (collapse)**	4.0	1.4	8.0	50	15	300	-0.50	-4 to -10
	AWSI	**Arctic Winter Sea Ice (collapse)**	6.3	4.5	8.7	20	10	100	0.60	0.6 to 1.2
	EAIS	**East Antarctic Ice Sheet (collapse)**	7.5	5.0	10.0	?	10k	?	0.60	2.0
	REEF	**Low -latitude Coral Reefs (die-off)**	1.5	1.0	2.0	10	~	~	~	~
Regional impact tipping elements	**PFAT**	**Boreal Permafrost (abrupt thaw)**	1.5	1.0	2.3	200	100	300	Abrupt thaw adds 50% to gradual: 10 GtC/14 GtCe/ .04°C per °C @2100; 25 GtC/35 GtCe/ .11°C per °C @2300	~
	BARI	Barents Sea Ice (abrupt loss)	1.6	1.5	1.7	25	?	?	~	+
	GLCR	**Mountain Glaciers (loss)**	2.0	1.5	3.0	200	50	1k	0.08	+
	SAHL	**Sahel and W. African Monsoon (greening)**	2.8	2.0	3.5	50	10	500	~	+
	BORF	**Boreal Forest (southern dieback)**	4.0	1.4	5.0	100	50	?	+52GtC / net -0.18°C	-0.5 to -2
	TUND	**Boreal Forest (northern expansion)**	4.0	1.5	7.2	100	40	?	-6 GtC / net +0.14°C	0.5-1.0

*Feedback strength in °C per °C for abrupt permafrost thaw is calculated relative to preindustrial and declines with further degrees of warming (by ~21% per °C).

calving (*37*, *38*) and shows early warning signals consistent with approaching a tipping point in west Greenland (*11*). Both ice sheet modeling and paleoclimate data indicate that a GrIS tipping point can occur when the melt-elevation feedback gets strong enough to support self-propelling melt (as an ice sheet surface loses height, it enters warmer air and thus melts faster) (*1*). Different models give a critical threshold of ~1.6°C (0.8 to 3.2°C) (*39*), ~1.5°C (*40*), or 2.7±0.2°C (*41*). Paleoclimate and model evidence shows that ice only reaches full coverage below ~0.3 to 0.5°C [~300 parts per million (ppm) CO_2] (*39*, *42*). Hysteresis allows GrIS to exist above this growth threshold once formed (*39*) but paleorecords indicate that GrIS partially retreats above this threshold (*42*) and likely collapsed during the long MIS-11 interglacial which was considerably warmer (>1.5°C) (*43*). A coupled ice sheet–atmosphere model found no collapse threshold (*44*), leading AR6 to state that there is limited evidence for irreversible GrIS loss below 3°C (*21*). However, some irreversible loss occurs beyond 3.5 m sea level equivalent (equivalent to ~2 to 2.5°C) (*44*), indicating self-perpetuating feedback. GrIS collapse would shift the Earth system to a unipolar icehouse state and affect other tipping elements (particularly the AMOC), hence qualifying as a global core tipping element (high confidence). Our best estimates for GrIS include a threshold of ~1.5°C (0.8 to 3°C) (high confidence), timescales of 10 ky (1 to 15 ky) (medium confidence), and GMST feedback of ~+0.13°C (regional ~+0.5 to 3°C) (low confidence). The timescale of ice sheet meltdown gets shorter the greater the temperature threshold is exceeded (*40*), with a minimum of ~1000 years.

West Antarctic ice sheet (WAIS)

Large parts of the WAIS are grounded below sea level; if the grounding line in these marine ice sheet basins reaches retrograde slopes, this may lead to the onset of marine ice sheet instability (MISI) and crossing of a tipping point (*7*, *8*, *45*). MISI is based on a feedback between the grounding line retreat and ice flux across the grounding line as it reaches thicker ice. This can lead to self-sustaining retreat and is hypothesized to have driven past collapses of the WAIS during previous warmer interglacial periods with high sea levels (*21*, *46*). Some glaciers in the Amundsen Sea Embayment are already close to this threshold and experiencing substantial grounding line retreat (*9*). The grounding line of Thwaites glacier is only ~30 km away from the subglacial ridge and retreating at ~1 km per year (*47*); eventual collapse may already be inevitable (*10*, *45*). Models support irreversible retreat being underway for present levels of ocean warming (*25*, *48*) and suggest that losing Thwaites glacier may destabilize much of WAIS (*7*). Under sustained

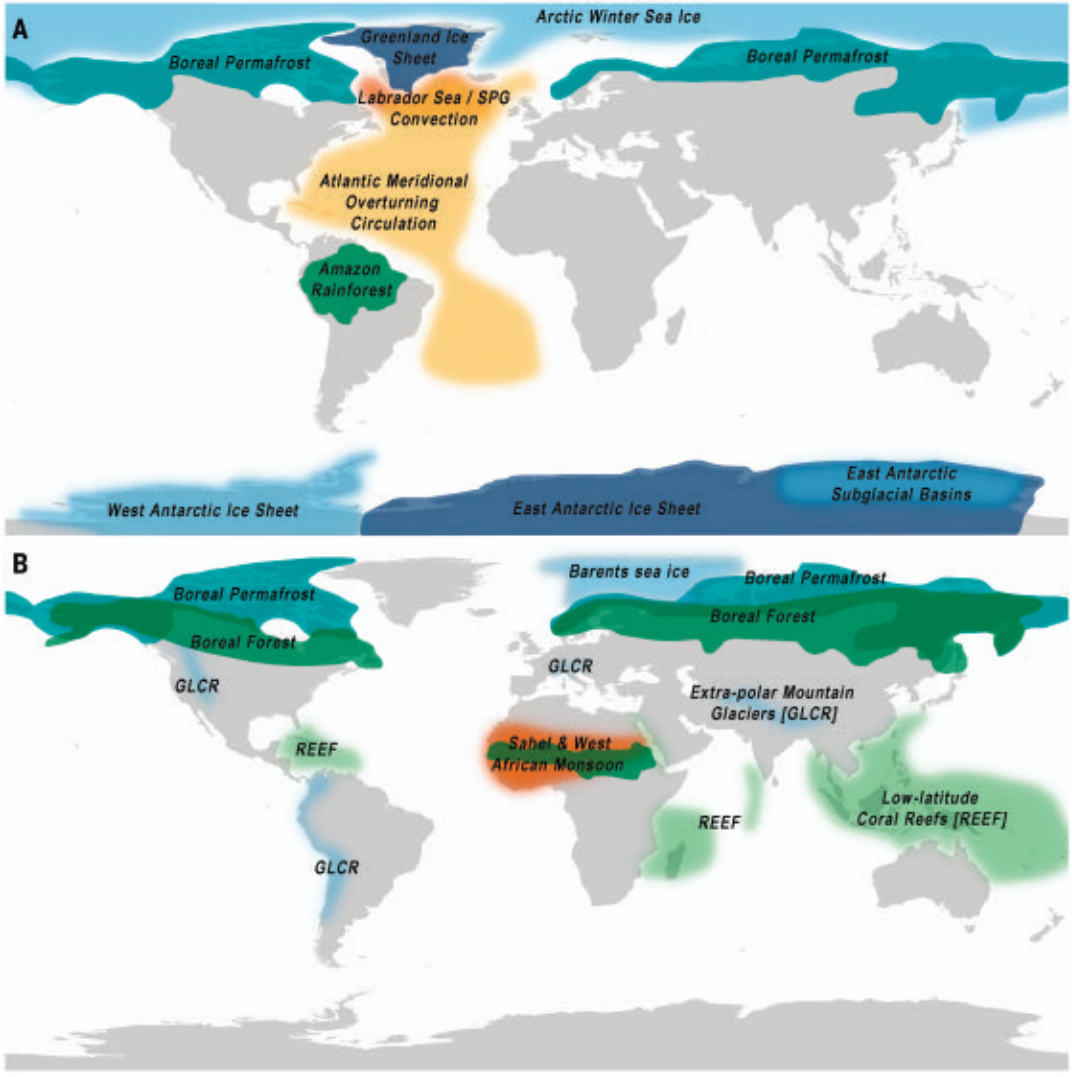


Fig. 1. Maps showing the global core (**A**) and regional impact (**B**) climate tipping elements identified in this study. Blue, green, and orange areas represent cryosphere, biosphere, and ocean-atmosphere elements, respectively.

1°C warming one model shows partial WAIS collapse with mass loss peaking at ~2°C warming (*25*). Hence we retain WAIS as a core global tipping element (high confidence) with a best estimate threshold of ~1.5°C [1 to 3°C, down from 3.5 to 5.5°C in (*1*)] (high confidence), timescales of 2 ky (500 years to 13 ky) (medium confidence), and GMST feedback of ~+0.05°C (regional ~+1°C) (low confidence). Higher threshold exceedance reduces the transition timescale to a minimum of ~500 years (*40*).

East Antarctic subglacial basins (EASB)

Recent data and models have shown that several subglacial basins of the EAIS—particularly the Wilkes, Aurora, and Recovery Basins—are also affected by MISI (*21*, *25*, *49*–*51*). These basins may also be subject to marine ice cliff instability in which the collapse of floating ice shelves creates unstable ice cliffs at the marine edge of the ice sheet which can retreat faster, but the importance of this process is disputed (*49*, *51*). One model indicates that Wilkes collapse may be committed by 3 to 4°C (*25*). Paleoclimate evidence for a higher mid-Pliocene sea level (+5 to 25 m) indicates that parts of the EASB (together with the GrIS and WAIS) were likely absent when the world was ~2.5 to 4°C warmer (*21*, *42*, *52*). By contrast sea levels of +6 to 13 m at 1.1 to 2.1°C in MIS-11 do not require substantial EASB contribution (assuming WAIS and GrIS were lost) (*50*). Hence we class EASB as a core global tipping element (high confidence) with best estimates for a tipping threshold of 3°C (2 to 6°C) (medium confidence), timescales of 2 ky (500 y to 10 ky) (medium confidence), and an uncertain GMST feedback provisionally assumed to be similar to WAIS (i.e., ~+0.05°C) (low confidence).

East Antarctic ice sheet (EAIS)

The land-grounded bulk of the EAIS is the world's largest ice sheet, containing the equivalent of ~50 m of sea level potential (*25*). Paleorecords indicate that growth occurred once atmospheric CO_2 fell below ~650 to 1000 ppm (~6 to 9°C) (*42*). Modeled ice sheets often exhibit alternative ice-covered or ice-free stable states for a range of global boundary conditions (*53*). As a result of this hysteresis the EAIS is expected to remain stable at CO_2 levels well beyond 650 ppm, helping it to survive through the warm mid-Miocene Climatic Optimum ~16 Mya (~2 to 4°C) (*42*). However, long-term stabilization at >1000 ppm CO_2 and ~8 to 10°C warming could cause total disintegration (*25*). Once past this threshold, self-perpetuating feedbacks amplify ice loss (*39*). The loss of EAIS would have global effects and hence is categorized as a global core

tipping element (medium confidence). Although unlikely, under high emissions [e.g., Representative Concentration Pathway (RCP) 8.5] and high climate sensitivity it might conceivably be committed to during this century or thereafter. Our best estimates for the EAIS are a tipping threshold of ~7.5°C (5 to 10°C) (medium confidence), timescales of >10 ky (medium confidence), and GMST feedback of ~+0.6°C (regional ~+2°C) (low confidence).

Boreal permafrost (PFTP/PFAT)

Permanently frozen soils and sediments in boreal regions contain ~1035 gigatonnes of carbon (GtC) that can be partly released as CO_2 and methane upon thawing (*54*). Although initially lacking evidence for a synchronous large-scale threshold (*1*), subsequent assessments recognized that part(s) of the permafrost could be considered a tipping element (*3*, *17*). Here we separate permafrost into three components with different dynamics: gradual thaw [PFGT; a threshold-free feedback (high confidence)] (*54*–*56*) (see SM); abrupt thaw [PFAT; a regional impact tipping element (medium confidence)], and collapse [PFTP: a global core tipping element (low confidence)]. Abrupt thaw processes (PFAT) such as slope slumping and thermokarst lake formation (*54*) could increase emissions by 50 to 100% relative to gradual thaw (*57*), involve localized tipping dynamics (e.g., continued thaw subsidence after initiation) and could occur near-synchronously on a subcontinental scale. Our best estimates for PFAT include a tipping threshold of 1.5°C (1 to 2.3°C) (medium confidence), a timescale of 200 years (100 to 300<years) (medium confidence), and an additional ~50% emissions beyond gradual thaw (~10 to 25 GtC per °C) (medium confidence). Finally, abrupt permafrost drying at ~4°C (*58*) and/or sufficiently rapid regional warming (>9°C) corresponding to ~5°C globally (*17*, *59*) could act as a trigger for permafrost collapse (PFTP) driven by internal heat production in carbon-rich permafrost—"the compost bomb" instability (*60*, *61*). The Yedoma deep ice- and carbon-rich permafrost (containing ~115 GtC in Yedoma deposits, ~400 GtC across the Yedoma domain) is particularly vulnerable as fast thaw processes can expose previously isolated deep deposits (*54*, *59*). This and other carbon-rich regions vulnerable to abrupt drying at >4°C (*58*) could have considerable feedback to global temperatures. Our best estimates for PFTP include a threshold of 4°C (3 to 6°C) (low confidence), a timescale of 50 years (10 to 300 years) (medium confidence), and emissions on the order of ~125 to 250 GtC (ΔGMST ~+0.2 to 0.4°C) (low confidence).

Mountain glaciers (GLCR)

Alpine glaciers outside Greenland and Antarctica have individual mass balance thresholds and elevation feedbacks yet large-scale synchronous losses are projected in several key regions at specific global warming levels. In transient simulations, peak water from European glacier melt is expected at ~1°C (*62*) with near-total loss expected to be committed at ~2°C (*20*). Global peak water occurs at ~2°C but committed eventual loss is expected at lower temperatures (*63*). Long model integrations show that global warming of 1.5 to 2°C is sufficient to lead to the eventual loss of most extra-polar glaciers (and possibly even polar glaciers) (*40*, *64*). RCP4.5 (>2°C by 2100) puts most lower-latitude glaciers on a path to significant losses beyond 2100 (*21*). Glaciers in High Mountain Asia last longer than elsewhere but reach peak water at ~2°C with significant social impacts for South Asia (*62*). Given the considerable human impacts of glacier loss (*63*) we categorize mountain glaciers as a regional impact tipping element (medium confidence). Our best estimate includes a threshold of ~2°C (1.5 to 3°C) (medium confidence), a timescale of 200 years (50 years to 1 ky) (medium confidence), and GMST feedback of ~+0.08°C (regionally greater) (low confidence).

Southern Ocean sea ice features abrupt events in some climate models (*19*) but because of uncertain dynamics and low confidence in projections it is classed as an uncertain tipping element (see SM). Marine methane hydrates are classed as a threshold-free feedback and Tibetan plateau snow is classed as uncertain (see SM).

Ocean-Atmosphere (circulation)

North Atlantic subpolar gyre / Labrador-Irminger Sea convection (LABC)

Convection in the Labrador and Irminger Seas in the North Atlantic—part of the subpolar gyre (SPG)—abruptly collapses in some models as a result of warming-induced stratification, a state which is then maintained by self-reinforcing convection feedbacks (*19*, *65*) giving two alternative stable states with or without deep convection. Abrupt future SPG collapse occurs in nine runs across five CMIP5 models at 1.1 to 2.0°C, in one additional model run at 3.8°C (*19*, *65*), and in four CMIP6 models in the 2040s (~1 to 2°C) (*66*). In some models SPG collapse affects AMOC strength but SPG and AMOC have distinct feedback dynamics and patterns of impacts, and SPG collapse can occur much faster than AMOC collapse. The North Atlantic cooling trend (i.e., the "warming hole") is centered over the SPG and in models is often closely linked to SPG weakening (*65*, *66*), although others have associated it with AMOC slowdown (*67*). SPG collapse causes a concentrated North Atlantic regional cooling of ~2 to 3°C, potential global cooling of up to ~0.5°C, a northward-shifted jet stream, weather extremes in Europe, and southward shift of the intertropical convergence zone (ITCZ) (*65*, *66*). Given clear tipping dynamics and global impact we class SPG as a global core tipping element (medium confidence), with a best estimate threshold of ~1.8°C (1.1 to 3.8°C) (high confidence), a timescale of 10 years (5 to 50 years) (high confidence), and GMST feedback of ~0.5°C (regional ~−3°C) (low confidence).

Atlantic meridional overturning circulation (AMOC)

The AMOC is self-sustaining due to salt-advection feedback (northward movement of warm water increases its density as a result of cooling and evaporation supporting the deep convection that drives the circulation). Import of salt at the southern boundary of the Atlantic also supports alternative strong and weak AMOC stable states with multiple abrupt switches between them observed in the past (*68*). Global warming increases Arctic precipitation, freshwater runoff from Greenland, and sea surface temperatures, which together slow down the AMOC by inhibiting deep convection. The AMOC is inferred from some reconstructions to have weakened by ~15% over the past ~50 years (*67*) and early warning signals in indirect AMOC footprints are consistent with the current AMOC "strong" state losing stability (*12*). However, the IPCC gives low confidence on historical AMOC trends (*21*). The Special Report on the Ocean and Cryosphere in a Changing Climate (SROCC) (*22*) assessed AMOC collapse occurring during the 21st century to be very unlikely but physically plausible; however, this was increased to unlikely (medium confidence) in AR6 (*21*). AMOC collapse is triggered in three runs of one CMIP5 model at 1.4 to 1.9°C and in two runs of an additional model at 2.2 to 2.5°C (*19*, *65*) in contrast to gradual declines in other CMIP5 and CMIP6 models (*21*). However, AR6 assessed CMIP models as generally tending toward "unrealistic stability" with respect to observational constraints (*14*, *21*). They also neglect meltwater forcing from rapid GrIS melt (*21*, *69*). Both factors make the AMOC more vulnerable to collapse. AMOC collapse would have global impacts on temperature and precipitation patterns including North Atlantic cooling, Southern Hemisphere warming, southward-shifted ITCZ, monsoon weakening in Africa and Asia, monsoon strengthening in the Southern Hemisphere leading to drying in the Sahel and parts of Amazonia, and reduced natural carbon sinks (*70*–*73*). Hence AMOC is retained as a core global tipping element (medium confidence) with a best estimate threshold of ~4°C [1.4 to 8°C versus 3.5 to 5.5°C in (*1*)] (low confidence), timescales of ~50 years (15 to 300 years) (medium confidence), and a GMST feedback of −0.5°C (low confidence) [regional −4 to −10°C, highly heterogeneous global pattern (medium confidence)].

The Indian summer monsoon (and other monsoon systems) have been reclassified as uncertain climate tipping elements because of a lack of evidence for a warming-related threshold behavior. Equatorial stratocumulus cloud breakup and Indian Ocean upwelling are uncertain as a result of limited evidence. Global ocean anoxia is uncertain because the global warming level required for weathering-induced anoxia is unclear. ENSO is reclassified as an unlikely tipping element as it lacks a clear self-perpetuation threshold. Arctic ozone hole expansion is reclassified as unlikely as it is now improbable that it would be triggerable as a result of climate change. The Northern Polar Jet stream is classed unlikely because instability as a result of climate change remains uncertain and no threshold has been proposed. (All of the above elements are discussed in more detail in the supplementary materials.)

Biosphere

Amazon rainforest (AMAZ)

The Amazon forest biome stores ~150 to 200 GtC (*3*, *74*, *75*) and historically has been a important sink for human CO_2 emissions (*15*). However, in intact forests this sink has declined since the 1990s (*15*) and ~17% of the Amazon forest has been lost to deforestation since the 1970s, a rate that has accelerated since 2019 (*75*). A combination of a climate change-induced drying trend, unprecedented droughts, and anthropogenic degradation in the south and east has led to the biome as a whole becoming a net carbon source (*74*). It has also lost resilience across ~76% of its area (*13*). Rainfall is projected to further decline and the dry season is expected to lengthen in southern and eastern areas of the forest in response to further warming, likely worsening this trend (*75*). The Amazon forest recycles around a third of the Amazon basin's rainfall on average (*76*) and up to ~70% in parts of the basin (*77*), particularly during the critical dry season as the forest maintains transpiration fluxes (*76*). This and localized fire feedbacks mean that ~40% of the Amazon forest is estimated to currently be in a bistable state, increasing to ~66% on an RCP8.5 trajectory (*18*, *77*), and rainforest loss could initiate self-reinforcing drying that tips this portion into a degraded or savanna-like state. Widespread Amazon dieback was originally projected at either 3 to 4°C of warming or ~40% deforestation (*78*) but uncertain synergistic interaction might lower the deforestation threshold to only ~20 to 25% (*79*). More recent ESMs tend not to simulate climate-induced Amazon dieback and emergent constraints indicate lower rainforest sensitivity to warming (*80*). However, two CMIP5 models exhibit dieback at 2.5 and 6.2°C (*19*). Additionally, CMIP5 ESMs underestimate observed tree mortality (*15*) and likely overestimate CO_2 fertilization (*81*), potentially making these models undersensitive to dieback. Given the size of the region affected by even partial dieback and its global impacts we categorize the Amazon forest as a global core tipping element (medium confidence). Our best estimates for AMAZ are a threshold of ~3.5°C (2 to 6°C) independent of deforestation (likely lower with deforestation) (low confidence), timescales of 100 years (50 to 200 years) (low confidence), and partial dieback of 40% (i.e., current bistable area) leading to emissions of ~30 GtC along with biogeophysical feedbacks (see SM) to a GMST feedback of ~+0.1° (regional +0.4 to 2°C) (medium confidence).

Boreal forest (BORF/TUND)

The boreal forest (or taiga) encircling the Arctic region features multiple stable states of tree cover as a result of feedbacks including albedo and fire (*82*, *83*). We classify it as a regional impact tipping element with two potential CTPs associated with abrupt dieback at its southern edge (BORF) (medium confidence) and abrupt expansion at its northern edge (tundra greening) (TUND) (medium confidence). Warming is projected to destabilize the southern edge, where factors such as hydrological changes, increased fire frequency, and bark beetle outbreaks can lead to self-reinforcing feedbacks driving regionally synchronized forest dieback (on the order of 100 km) to a grass-dominated steppe or prairie state (*83*). Models project that regime shifts may start in this area at ~1.5°C and become widespread by >3.5°C (*84*, *85*). Dieback may also be rate-dependent (*85*). Our best estimates for BORF is a threshold of ~4°C (1.4 to 5°C) (low confidence), timescales of 100 years (50+ years) (low confidence), and partial (~50%) dieback leading to emissions of ~52 GtC, which—along with countervailing biogeophysical feedbacks such as increased albedo and reduced evapotranspiration—leads to a net GMST feedback of ~−0.18°C (medium confidence) [regional ~−0.5 to 2°C (low confidence)]. Northward expansion of the forest into the current tundra biome may also feature self-perpetuation dynamics (e.g., by causing further local warming through albedo feedback). Models suggest that regime shifts may begin in this northern area at ~1.5°C and become widespread by ~3.5°C (*84*), with abrupt high latitude forest expansion occurring in one CMIP5 model at 7.2°C (*19*). For TUND our best estimates are for a threshold at ~4°C (1.5 to 7.2°C) (low confidence), timescales of 100 years (40+ years) (low confidence), and partial (~50%) uptake of ~6 GtC which along with countervailing biogeophysical feedbacks (decreased albedo, increased evapotranspiration) leads to a net GMST feedback of +0.14°C per °C (regional ~+0.5 to 1°C) (low confidence).

Sahel vegetation and the West African monsoon (SAHL)

Paleoevidence indicates multiple abrupt shifts into and out of African Humid Periods with associated greening of the Sahara in response to gradual changes in orbital forcing (*86*). AMOC weakening and associated warming of the Equatorial East Atlantic also caused past collapses of the West African monsoon (WAM) (*70*, *86*, *87*). Dust aerosol-rainfall positive feedbacks amplify change alongside well-established vegetation-rainfall positive feedbacks but many models still underestimate self-amplifying feedbacks and cannot reproduce the extent of past rainfall and vegetation changes (*86*). By contrast, a model optimized against present observations and mid-Holocene reconstructions recently reproduced abrupt transitions in Saharan vegetation with potential tipping dynamics (*88*). In future projections with GHG forcing, global (CMIP5 and CMIP6) and some regional Coordinated Regional Climate Downscaling Experiment (CORDEX) climate models tend to predict strengthening of the WAM and wetting and northward expansion of the central and eastern Sahel (as well as drying in coastal west Africa) (*23*, *70*, *89*–*91*), which tend to green the Sahel (*86*). Abrupt increases in vegetation in the Eastern Sahel occur in three ESM runs at 2.1 to 3.5°C (*19*). In other global models more gradual WAM strengthening and vegetation shifts are predicted but in some regional climate models the WAM instead collapses (*89*). Clearly the existence of a future tipping threshold for the WAM and Sahel remains uncertain as does its sign but given multiple past abrupt shifts, known weaknesses in current models, and huge regional impacts but modest global climate feedback, we retain the Sahel/WAM as a potential regional impact tipping element (low confidence). We adopt the scenario of abrupt wetting and greening with a threshold of ~2.8°C (2 to 3.5°C) (low confidence), a timescale of 50 years (10 to 500 years) (low confidence), and uncertain Earth system impacts (regional warming) (medium confidence).

Low-latitude coral reefs (REEF)

Tropical and subtropical coral reefs are threatened by anthropogenic pressures including overfishing, direct damage, sedimentation, ocean acidification, and global warming (*92*). When water temperatures exceed a certain threshold coral irreversibly expel their symbiotic algae resulting in coral bleaching, thereby triggering coral death (*93*). Ocean acidification worsens warming-induced degradation. Coral collapse would remove one of the Earth's most biodiverse ecosystems, affecting the wider marine food web, ocean nutrient and carbon cycling, and livelihoods of millions of people worldwide (*92*). Although coral bleaching is a

localized process synchronous bleaching can occur at the ~1000 km scale (as seen for the Great Barrier Reef), and further warming is expected to cause widespread bleaching (*93*). Adaptation may be possible with slower warming rates (*92*) but the IPCC has projected 70 to 90% tropical and subtropical coral reef loss at 1.5°C with near total loss by 2°C (*90*). Given regionally synchronized tipping dynamics with significant human but indirect climate impacts, we categorize warm-water coral reefs as a regional impact tipping element (high confidence). Our best estimates are for a threshold of ~1.5°C (1 to 2°C) (high confidence), timescales of ~10 years (medium confidence), and negligible GMST feedback (high confidence). Beyond the biosphere elements listed, the ocean biological pump and land/ocean carbon sink are unlikely to be tipping elements although they may feature nonlinearities (see SM).

Implications for climate policy and preventing dangerous levels of global warming

Figure 2A summarizes our temperature threshold estimates for each tipping element making the shortlists (others are summarized in the supplementary text). Here we define crossing a CTP as "possible" beyond its minimum temperature threshold and "likely" beyond its central estimate.

This revised assessment of CTPs has significant implications for climate policy by determining levels of global warming that prevent tipping to either committed changes in Earth system function or major damage to future societies. A risk minimization approach such as this seeks to avoid minimum estimated thresholds but this no longer appears possible for some tipping elements.

Current warming is ~1.1°C above preindustrial and even with rapid emission cuts warming will reach ~1.5°C by the 2030s (*23*). We cannot rule out that WAIS and GrIS tipping points have already been passed (see above) and several other tipping elements have minimum threshold values within the 1.1 to 1.5°C range. Our best estimate thresholds for GrIS, WAIS, REEF, and abrupt permafrost thaw (PFAT) are ~1.5°C although WAIS and GrIS collapse may still be avoidable if GMST returns below 1.5°C within an uncertain overshoot time (likely decades) (*94*). Setting aside achievability (and recognizing internal climate variability of ~±0.1°C), this suggests that ~1°C is a level of global warming that minimizes the likelihood of crossing CTPs. This is consistent with the <0.5 to 1°C range of Holocene temperature variability whereas past interglacials ≤1.5 to 2°C had up to 10 to 13 m higher sea level (*21*, *95*).

The chance of triggering CTPs is already non-negligible and will grow even with stringent climate mitigation (SSP1-1.9 in Fig. 2, B and C). Nevertheless, achieving the Paris Agreement's aim to pursue efforts to limit warming to 1.5°C would clearly be safer than keeping global warming below 2°C (*90*) (Fig. 2). Going from 1.5 to 2°C increases the likelihood of committing to WAIS and GrIS collapse near complete warm-water coral die-off, and abrupt permafrost thaw; further, the best estimate threshold for LABC collapse is crossed. The likelihood of triggering AMOC collapse, Boreal forest shifts, and extra-polar glacier loss becomes non-negligible at >1.5°C and glacier loss becomes likely by ~2°C. A cluster of abrupt shifts occur in ESMs at 1.5 to 2°C (*19*). Although not tipping elements, ASSI loss could become regular by 2°C, gradual permafrost thaw would likely become widespread beyond 1.5°C, and land carbon sink weakening would become significant by 2°C.

Recent net zero targets if implemented could limit peak warming to ~1.95°C (1.3 to 3°C), but as of November 2021 current policies are estimated to result in ~2.6 °C (1.9 °C to 3.7 °C) by 2100 (*96*). Therefore 2 to 3°C by 2100 is currently likely with matching of pledges with policies key to determining where warming ends up in this range. Going from 2 to 3°C, maximum estimated thresholds for abrupt permafrost thaw, GrIS, WAIS, and extra-polar glaciers are passed, suggesting that tipping them would become very likely. The likelihood of triggering EASB collapse, Amazon dieback, and West African monsoon shift (Sahel greening) becomes non-negligible at ~2°C and increases at ~3°C. Subpolar gyre collapse, boreal forest dieback, and AMOC collapse also become more likely. Although not tipping elements, above 2°C the Arctic would very likely become summer ice–free and land carbon sink-to-source transitions would become widespread.

If the moderate ambition of current policies is not improved and climate sensitivity or carbon cycle feedbacks turn out to be higher than the median assumption then warming of up to ~4°C is possible by 2100, and >4°C cannot be ruled out if future policy ambition declines and/or implementation falters. Going from 3 to 5°C, EASB collapse becomes very likely, Amazon dieback becomes likely >3.5°C, boreal forest shifts likely >4°C, and large-scale permafrost collapse becomes possible at >3°C and likely >4°C. AMOC collapse may become likely >4°C but with high uncertainty (1.4 to 8°C range) and AWSI collapse becomes possible >4.5°C. Warming of >5°C, although very unlikely this century, becomes plausible in the longer-term under higher climate sensitivities with current or reversed policies. This risks EAIS collapse and a commitment of ~55 m of sea level rise if warming stabilizes >5°C for multiple centuries. Other tipping elements, if not already triggered—e.g., Amazon dieback, widespread Permafrost collapse—would very likely be committed and AMOC collapse and AWSI collapse would become increasingly likely. Equatorial stratocumulus cloud breakup occurs in one model beyond ~6°C (*97*) and if plausible would represent a global CTP to a "hothouse" climate state (*3*).

Discussion

Tipping elements and their tipping points were treated independently in this assessment but there are multiple causal interactions between them with risks of triggering cascades among CTPs (*2*, *4*, *16*), some mediated through temperature. The strength and in some cases even the sign of identified interactions is uncertain (*4*). Nevertheless, their combined effect tends to lower CTP temperature thresholds (*6*, *16*). The present assessment would likely amplify this effect, further strengthening the incentive for ambitious mitigation.

Some of the threshold and impact estimates are highly uncertain (e.g., AMOC, BORF/TUND, AMAZ, SAHL, PFTP) as are the transition timescales of many elements. Some proposed elements remain too uncertain to categorize (e.g., EQSC, ANOX, INSM, AABW, Congo rainforest), and others considered unlikely to feature tipping dynamics (e.g., ENSO, JETS) cannot yet be fully ruled out (see SM). Other tipping elements may yet be discovered. It may be possible to safely overshoot CTPs in slower elements such as ice sheets (*94*) but the allowable overshoot times need further research. Spatial pattern formation might allow some biosphere elements to evade direct tipping (*98*) but this needs to be assessed.

To further our understanding of the likelihood of crossing CTPs an updated expert elicitation [building on (*4*)] is overdue. A horizon-scanning exercise and systematic scanning of CMIP6 model output [following (*19*)] and a tipping point model intercomparison project could help identify more candidate tipping elements and refine assessment of their likelihood. Further model improvements and model-data intercomparison are essential. Early warning methods are starting to reveal whether tipping elements are destabilizing for parts of GrIS (*11*), AMOC (*12*), and the Amazon (*13*), and can reveal proximity to a CTP (*11*). These could be augmented with deep learning techniques (*99*). Systematic application to observational and remotely sensed data together with targeted new observing systems could begin to provide a CTP early warning system (*100*).

Conclusion

The UNFCCC stipulates that all countries commit to avoid dangerous climate change, translated through the Paris climate agreement into keeping GMST well below 2°C and aiming for 1.5°C. Our assessment of climate tipping elements and their tipping points suggests that danger may be approached

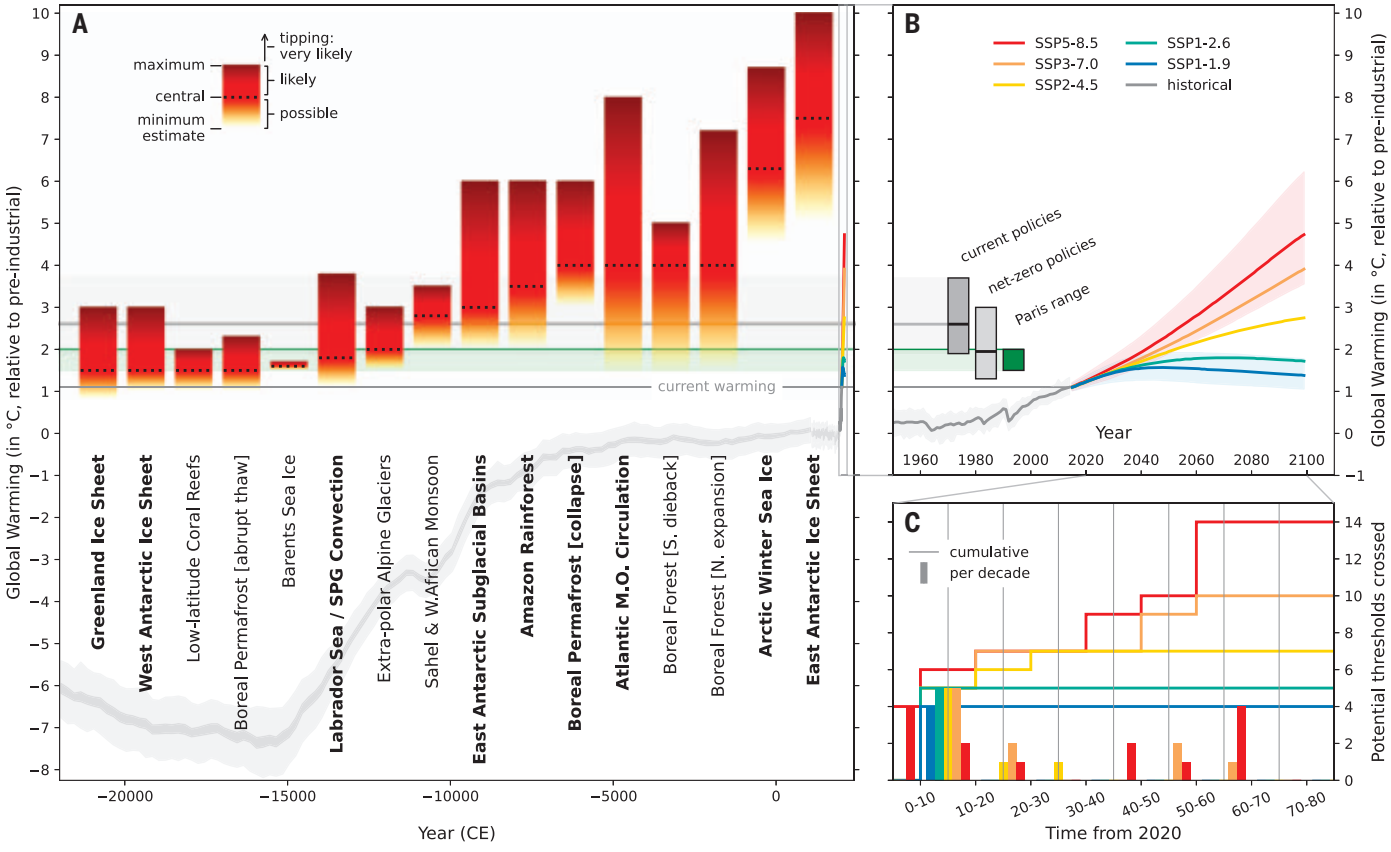


Fig. 2. Our global warming threshold estimates for global core and regional impact climate tipping elements. Tipping elements (**A**) relative to IPCC SSP projections and likely future scenarios given current policies and targets (**B**) and how many thresholds may be crossed per Shared Socioeconomic Pathways (SSP) projection (**C**). Bars in (A) show the minimum (base, yellow), central (line, red), and maximum (top, dark red) threshold estimates for each element (bold font, global core; regular font, regional impact), with a paleorecord of GMST over the past ~25 ky (*95*) and projections of future climate change (green, SSP1-1.9; yellow, SSP1-2.6; orange, SSP2-4.5; red, SSP3-7.0; purple, SSP5-8.5) from IPCC AR6 (*23*) shown for context. Future projections are shown in more detail in (B) along with estimated 21st century warming trajectories for current and net-zero policies (gray bars, extending into (A); horizontal lines show central estimates, bar height the uncertainty ranges) as of November 2021 (*96*) versus the Paris Agreement range of 1.5 to <2°C (green bar). The number of thresholds potentially passed in the coming decades depending on SSP trajectory in (C) is shown per decade (bars) and cumulatively (lines).

even earlier. The Earth may have left a safe climate state beyond 1°C global warming. A significant likelihood of passing multiple climate tipping points exists above ~1.5°C, particularly in major ice sheets. Tipping point likelihood increases further in the Paris range of 1.5 to <2°C warming. Current policies leading to ~2 to 3°C warming are unsafe because they would likely trigger multiple climate tipping points. Our updated assessment of climate tipping points provides strong scientific support for the Paris Agreement and associated efforts to limit global warming to 1.5°C.

Materials and Methods

We mined the literature subsequent to (*1*), including studies of paleoclimate change, observed change, early warning signals, future model projections, underlying theory, and existing assessments, to draw up a longlist of possible candidate tipping elements (table S3). For each we extracted information on evidence for self-perpetuation, temperature thresholds, hysteresis/irreversibility, transition timescales, and global/regional impacts on climate, on which we then use subjective expert judgment to determine our best estimates. From this evidence (or lack of it) we drew up shortlists (Table 1) of 'core' global tipping elements and regional 'impact' tipping elements (Fig. 1), for which we summarize the rationale in the main text and supplementary text S2 and S3. Candidates that did not make the shortlists (table S3) are classed as: (a) 'uncertain' tipping elements—due to limited evidence for self-perpetuating feedback and threshold behavior; (b) 'unlikely' tipping elements—possessing only localized tipping or nonfeedback response to climate change; and (c) 'threshold-free feedbacks'—where positive feedbacks exist but are not strong enough to self-perpetuate. Different parts or phenomena of some systems—notably permafrost—are assigned to different categories. We give (very low, low, medium, high, very high) confidence levels based on the IPCC's confidence rating system (as a product of the authors' judgements of both the robustness and the degree of agreement of the assessed literature) (*101*) for the estimates of central, minimum, and maximum temperature thresholds, timescales of transition, and global and local impacts on climate (supplementary text S2). We define crossing a CTP as 'possible' beyond its minimum temperature threshold and 'likely' beyond its best estimate. Differences with past lists of tipping elements are described in table S4.

REFERENCES AND NOTES

1. T. M. Lenton *et al.*, Tipping elements in the Earth's climate system. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1786–1793 (2008). doi: 10.1073/pnas.0705414105; pmid: 18258748
2. T. M. Lenton *et al.*, Climate tipping points - too risky to bet against. *Nature* **575**, 592–595 (2019). doi: 10.1038/d41586-019-03595-0; pmid: 31776487
3. W. Steffen *et al.*, Trajectories of the Earth System in the Anthropocene. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8252–8259 (2018). doi: 10.1073/pnas.1810141115; pmid: 30082409
4. E. Kriegler, J. W. Hall, H. Held, R. Dawson, H. J. Schellnhuber, Imprecise probability assessment of tipping points in the climate system. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5041–5046 (2009). doi: 10.1073/pnas.0809117106; pmid: 19289827

5. H. J. Schellnhuber, S. Rahmstorf, R. Winkelmann, Why the right climate target was agreed in Paris. *Nat. Clim. Chang.* **6**, 649–653 (2016). doi: 10.1038/nclimate3013
6. Y. Cai, T. M. Lenton, T. S. Lontzek, Risk of multiple interacting tipping points should encourage rapid CO2 emission reduction. *Nat. Clim. Chang.* **6**, 520–525 (2016). doi: 10.1038/nclimate2964
7. J. Feldmann, A. Levermann, Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14191–14196 (2015). doi: 10.1073/pnas.1512482112; pmid: 26578762
8. M. S. Waibel, C. L. Hulbe, C. S. Jackson, D. F. Martin, Rate of Mass Loss Across the Instability Threshold for Thwaites Glacier Determines Rate of Mass Loss for Entire Basin. *Geophys. Res. Lett.* **45**, 809–816 (2018). doi: 10.1002/2017GL076470
9. E. Rignot, J. Mouginot, M. Morlighem, H. Seroussi, B. Scheuchl, Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. *Geophys. Res. Lett.* **41**, 3502–3509 (2014). doi: 10.1002/2014GL060140
10. I. Joughin, B. E. Smith, B. Medley, Marine Ice Sheet Collapse Potentially Under Way for the Thwaites Glacier Basin, West Antarctica. *Science* **344**, 735–738 (2014). doi: 10.1002/2014GL060140
11. N. Boers, M. Rypdal, Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024192118 (2021). doi: 10.1073/pnas.2024192118; pmid: 34001613
12. N. Boers, Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation. *Nat. Clim. Chang.* **11**, 680–688 (2021). doi: 10.1038/s41558-021-01097-4
13. C. A. Boulton, T. M. Lenton, N. Boers, Pronounced loss of Amazon rainforest resilience since the early 2000s. *Nat. Clim. Chang.* **12**, 271–278 (2022). doi: 10.1038/s41558-022-01287-8
14. W. Liu, S.-P. Xie, Z. Liu, J. Zhu, Overlooked possibility of a collapsed Atlantic Meridional Overturning Circulation in warming climate. *Sci. Adv.* **3**, e1601666 (2017). doi: 10.1126/sciadv.1601666; pmid: 28070560
15. W. Hubau *et al*., Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* **579**, 80–87 (2020). doi: 10.1038/s41586-020-2035-0; pmid: 32132693
16. N. Wunderling, J. F. Donges, J. Kurths, R. Winkelmann, Interacting tipping elements increase risk of climate domino effects under global warming. *Earth Syst. Dyn.* **12**, 601–619 (2021). doi: 10.5194/esd-12-601-2021
17. T. M. Lenton, Arctic climate tipping points. *Ambio* **41**, 10–22 (2012). doi: 10.1007/s13280-011-0221-x; pmid: 22270703
18. A. Staal *et al*., Hysteresis of tropical forests in the 21st century. *Nat. Commun.* **11**, 4978 (2020). doi: 10.1038/s41467-020-18728-7; pmid: 33020475
19. S. Drijfhout *et al*., Catalogue of abrupt shifts in Intergovernmental Panel on Climate Change climate models. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5777–E5786 (2015). doi: 10.1073/pnas.1511451112; pmid: 26460042
20. A. Levermann *et al*., Potential climatic transitions with profound impact on Europe. *Clim. Change* **110**, 845–878 (2012). doi: 10.1007/s10584-011-0126-5
21. B. Fox-Kemper *et al*., "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" V. Masson-Delmotte *et al*, Eds. (Cambridge Univ. Press, 2021).
22. M. Collins *et al*., "IPCC Special Report on the Ocean and Cryosphere in a Changing Climate" H.-O. Pörtner *et al*., Eds. (2019); pp. 589–655.
23. J. Y. Lee *et al*., "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021).
24. W. R. Boos, T. Storelvmo, Near-linear response of mean monsoon strength to a broad range of radiative forcings. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1510–1515 (2016). doi: 10.1073/pnas.1517143113; pmid: 26811462
25. J. Garbe, T. Albrecht, A. Levermann, J. F. Donges, R. Winkelmann, The hysteresis of the Antarctic Ice Sheet. *Nature* **585**, 538–544 (2020). doi: 10.1038/s41586-020-2727-5; pmid: 32968257
26. D. Chen *et al*., "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021).
27. M. Scheffer *et al*., Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009). doi: 10.1038/nature08227; pmid: 19727193
28. R. Ranasinghe *et al*., "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021), p. 73.
29. National Research Council, *Abrupt Climate Change: Inevitable Surprises* (The National Academies Press, 2002); https://www.nap.edu/catalog/10136.
30. R. E. Kopp, R. L. Shwom, G. Wagner, J. Yuan, Tipping elements and climate–economic shocks: Pathways toward integrated assessment. *Earths Future* **4**, 346–372 (2016). doi: 10.1002/2016EF000362
31. C. Hankel, E. Tziperman, The Role of Atmospheric Feedbacks in Abrupt Winter Arctic Sea Ice Loss in Future Warming Scenarios. *J. Clim.* **34**, 4435–4447 (2021). doi: 10.1175/JCLI-D-20-0558.1
32. P. J. Hezel, T. Fichefet, F. Massonnet, Modeled Arctic sea ice evolution through 2300 in CMIP5 extended RCPs. *Cryosphere* **8**, 1195–1204 (2014). doi: 10.5194/tc-8-1195-2014
33. D. Docquier, R. Fuentes-Franco, T. Koenigk, T. Fichefet, Sea Ice—Ocean Interactions in the Barents Sea Modeled at Different Resolutions. *Front. Earth Sci.* **8**, 1–21 (2020). doi: 10.3389/feart.2020.00172
34. F. Lehner, A. Born, C. C. Raible, T. F. Stocker, Amplified Inception of European Little Ice Age by Sea Ice–Ocean–Atmosphere Feedbacks. *J. Clim.* **26**, 7586–7602 (2013). doi: 10.1175/JCLI-D-12-00690.1
35. I. Allison *et al*., "The Copenhagen Diagnosis" (The Univ. of New South Wales Climate Change Research Centre, 2009); https://www.ccrc.unsw.edu.au/sites/default/files/Copenhagen_Diagnosis_LOW.pdf).
36. D. Notz, S. Community, Arctic Sea Ice in CMIP6. *Geophys. Res. Lett.* **47**, 1–26 (2020). doi: 10.1029/2019GL086749
37. M. D. King *et al*., Dynamic ice loss from the Greenland Ice Sheet driven by sustained glacier retreat. *Commun. Earth Environ.* **1**, 1–7 (2020). doi: 10.1038/s43247-020-0001-2
38. A. Shepherd *et al*., Mass balance of the Greenland Ice Sheet from 1992 to 2018. *Nature* **579**, 233–239 (2020).
39. A. Robinson, R. Calov, A. Ganopolski, Multistability and critical thresholds of the Greenland ice sheet. *Nat. Clim. Chang.* **2**, 429–432 (2012). doi: 10.1038/nclimate1449
40. J. Van Breedam, H. Goelzer, P. Huybrechts, Semi-equilibrated global sea-level change projections for the next 10 000 years. *Earth Syst. Dyn.* **11**, 953–976 (2020). doi: 10.5194/esd-11-953-2020
41. B. Noël, L. van Kampenhout, J. T. M. Lenaerts, W. J. van de Berg, M. R. van den Broeke, A 21st Century Warming Threshold for Sustained Greenland Ice Sheet Mass Loss. *Geophys. Res. Lett.* **48**, 1–9 (2021). doi: 10.1029/2020GL090471
42. G. L. Foster, E. J. Rohling, Relationship between sea level and climate forcing by CO2 on geological timescales. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1209–1214 (2013). doi: 10.1073/pnas.1216073110; pmid: 23292932
43. A. J. Christ *et al*., A multimillion-year-old record of Greenland vegetation and glacial history preserved in sediment beneath 1.4 km of ice at Camp Century. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021442118 (2021). doi: 10.1073/pnas.2021442118; pmid: 33723012
44. J. M. Gregory, S. E. George, R. S. Smith, Large and irreversible future decline of the Greenland ice sheet. *Cryosphere* **14**, 4299–4322 (2020). doi: 10.5194/tc-14-4299-2020
45. R. B. Alley *et al*., Oceanic Forcing of Ice-Sheet Retreat: West Antarctica and More. *Annu. Rev. Earth Planet. Sci.* **43**, 207–231 (2015). doi: 10.1146/annurev-earth-060614-105344
46. C. S. M. Turney *et al*., Early Last Interglacial ocean warming drove substantial ice mass loss from Antarctica. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 3996–4006 (2020). doi: 10.1073/pnas.1902469117; pmid: 32047039
47. H. Yu, E. Rignot, H. Seroussi, M. Morlighem, Impact of iceberg calving on the retreat of Thwaites Glacier, West Antarctica over the next century with different calving laws and ocean thermal forcing. *Geophys. Res. Lett.* **46**, 14539–14547 (2019). doi: 10.1002/2017GL072514
48. R. J. Arthern, C. R. Williams, The sensitivity of West Antarctica to the submarine melting feedback. *Geophys. Res. Lett.* **44**, 2352–2359 (2017). doi: 10.1002/2017GL072514
49. T. L. Edwards *et al*., Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature* **566**, 58–64 (2019). doi: 10.1038/s41586-019-0901-4; pmid: 30728522
50. P. U. Clark *et al*., Oceanic forcing of penultimate deglacial and last interglacial sea-level rise. *Nature* **577**, 660–664 (2020). doi: 10.1038/s41586-020-1931-7; pmid: 31996820
51. R. M. DeConto *et al*., The Paris Climate Agreement and future sea-level rise from Antarctica. *Nature* **593**, 83–89 (2021). doi: 10.1038/s41586-021-03427-0; pmid: 33953408
52. D. J. Wilson *et al*., Ice loss from the East Antarctic Ice Sheet during late Pleistocene interglacials. *Nature* **561**, 383–386 (2018). doi: 10.1038/s41586-018-0501-8; pmid: 30232420
53. E. Gasson, R. M. DeConto, D. Pollard, R. H. Levy, Dynamic Antarctic ice sheet during the early to mid-Miocene. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3459–3464 (2016). doi: 10.1073/pnas.1516130113; pmid: 26903645
54. E. A. G. Schuur *et al*., Climate change and the permafrost carbon feedback. *Nature* **520**, 171–179 (2015). doi: 10.1038/nature14338; pmid: 25855454
55. A. Vaks *et al*., Speleothems Reveal 500,000-Year History of Siberian Permafrost. *Science* **340**, 183–186 (2013). doi: 10.1038/nature14338; pmid: 25855454
56. J. G. Canadell *et al*., "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021).
57. M. R. Turetsky *et al*., Carbon release through abrupt permafrost thaw. *Nat. Geosci.* **13**, 138–143 (2020). doi: 10.1038/s41561-019-0526-0
58. B. Teufel, L. Sushama, Abrupt changes across the Arctic permafrost region endanger northern development. *Nat. Clim. Chang.* **9**, 858–862 (2019). doi: 10.1038/s41558-019-0614-6
59. J. Strauss *et al*., Deep Yedoma permafrost: A synthesis of depositional characteristics and carbon vulnerability. *Earth Sci. Rev.* **172**, 75–86 (2017). doi: 10.1016/j.earscirev.2017.07.007
60. C. M. Luke, P. M. Cox, Soil carbon and climate change: From the Jenkinson effect to the compost-bomb instability. *Eur. J. Soil Sci.* **62**, 5–12 (2011). doi: 10.1111/j.1365-2389.2010.01312.x
61. J. Hollesen, H. Matthiesen, A. B. Møller, B. Elberling, Permafrost thawing in organic Arctic soils accelerated by ground heat production. *Nat. Clim. Chang.* **5**, 574–578 (2015). doi: 10.1038/nclimate2590
62. M. Huss, R. Hock, Global-scale hydrological response to future glacier mass loss. *Nat. Clim. Chang.* **8**, 135–140 (2018). doi: 10.1038/s41558-017-0049-x
63. R. Hock *et al*., "IPCC Special Report on the Ocean and Cryosphere in a Changing Climate" H.-O. Pörtner *et al*., Eds. (Cambridge Univ. Press, 2019), pp. 131–202.
64. P. U. Clark *et al*., Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nat. Clim. Chang.* **6**, 360–369 (2016). doi: 10.1038/nclimate2923
65. G. Sgubin, D. Swingedouw, S. Drijfhout, Y. Mary, A. Bennabi, Abrupt cooling over the North Atlantic in modern climate models. *Nat. Commun.* **8**, 14375 (2017). doi: 10.1038/ncomms14375; pmid: 28198383
66. D. Swingedouw *et al*., On the risk of abrupt changes in the North Atlantic subpolar gyre in CMIP6 models. *Ann. N. Y. Acad. Sci.* **1504**, 187–201 (2021). doi: 10.1111/nyas.14659; pmid: 34212391
67. L. Caesar, S. Rahmstorf, A. Robinson, G. Feulner, V. Saba, Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature* **556**, 191–196 (2018). doi: 10.1038/s41586-018-0006-5; pmid: 29643485
68. J. Lynch-Stieglitz, The Atlantic Meridional Overturning Circulation and Abrupt Climate Change. *Annu. Rev. Mar. Sci.* **9**, 83–104 (2017). doi: 10.1146/annurev-marine-010816-060415; pmid: 27814029
69. J. Lohmann, P. D. Ditlevsen, Risk of tipping the overturning circulation due to increasing rates of ice melt. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017989118 (2021). doi: 10.1073/pnas.2017989118; pmid: 33619095
70. H. Douville *et al*., "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021), p. 73.
71. L. C. Jackson *et al*., Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM. *Clim. Dyn.* **45**, 3299–3316 (2015). doi: 10.1007/s00382-015-2540-2

72. S. Drijfhout, Competition between global warming and an abrupt collapse of the AMOC in Earth's energy imbalance. *Sci. Rep.* **5**, 14877 (2015). doi: 10.1038/srep14877; pmid: 26437599
73. A. Bozbiyik, M. Steinacher, F. Joos, T. F. Stocker, L. Menviel, Fingerprints of changes in the terrestrial carbon cycle in response to large reorganizations in ocean circulation. *Clim. Past* **7**, 319–338 (2011). doi: 10.5194/cp-7-319-2011
74. L. V. Gatti *et al*., Amazonia as a carbon source linked to deforestation and climate change. *Nature* **595**, 388–393 (2021). doi: 10.1038/s41586-021-03629-6; pmid: 34262208
75. Science Panel for the Amazon (SPA), "Executive Summary of the Amazon Assessment Report 2021" C. Nobre *et al*., Eds. (United Nations Sustainable Development Solutions Network, 2021); https://www.theamazonwewant.org/wp-content/uploads/2021/09/SPA-Executive-Summary-11Mb.pdf.
76. A. Staal *et al*., Forest-rainfall cascades buffer against drought across the Amazon. *Nat. Clim. Chang.* **8**, 539–543 (2018). doi: 10.1038/s41558-018-0177-y
77. D. C. Zemp *et al*., Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks. *Nat. Commun.* **8**, 14681 (2017). doi: 10.1038/ncomms14681; pmid: 28287104
78. C. A. Nobre *et al*., Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10759–10768 (2016). doi: 10.1073/pnas.1605516113; pmid: 27638214
79. T. E. Lovejoy, C. Nobre, Amazon Tipping Point. *Sci. Adv.* **4**, eaat2340 (2018). doi: 10.1126/sciadv.aat2340; pmid: 29492460
80. P. M. Cox *et al*., Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* **494**, 341–344 (2013). doi: 10.1038/nature11882; pmid: 23389447
81. C. Terrer *et al*., Nitrogen and phosphorus constrain the CO2 fertilization of global plant biomass. *Nat. Clim. Chang.* **9**, 684–689 (2019). doi: 10.1038/s41558-019-0545-2
82. B. Abis, V. Brovkin, Environmental conditions for alternative tree-cover states in high latitudes. *Biogeosciences* **14**, 511–527 (2017). doi: 10.5194/bg-14-511-2017
83. M. Scheffer, M. Hirota, M. Holmgren, E. H. Van Nes, F. S. Chapin3rd, Thresholds for boreal biome transitions. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21384–21389 (2012). doi: 10.1073/pnas.1219844110; pmid: 23236159
84. D. Gerten *et al*., Asynchronous exposure to global warming: Freshwater resources and terrestrial ecosystems. *Environ. Res. Lett.* **8**, 034032 (2013). doi: 10.1088/1748-9326/8/3/034032
85. C. D. Koven, Boreal carbon loss due to poleward shift in low-carbon ecosystems. *Nat. Geosci.* **6**, 452–456 (2013). doi: 10.1038/ngeo1801
86. F. S. R. Pausata *et al*., The Greening of the Sahara: Past Changes and Future Implications. *One Earth* **2**, 235–250 (2020). doi: 10.1016/j.oneear.2020.03.002
87. M. W. Schmidt, P. Chang, A. O. Parker, L. Ji, F. He, Deglacial Tropical Atlantic subsurface warming links ocean circulation variability to the West African Monsoon. *Sci. Rep.* **7**, 15390 (2017). doi: 10.1038/s41598-017-15637-6; pmid: 29133905
88. P. O. Hopcroft, P. J. Valdes, Paleoclimate-conditioning reveals a North Africa land-atmosphere tipping point. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2108783118 (2021). doi: 10.1073/pnas.2108783118; pmid: 34725155
89. A. Dosio *et al*., Projected future daily characteristics of African precipitation based on global (CMIP5, CMIP6) and regional (CORDEX, CORDEX-CORE) climate models. *Clim. Dyn.* **57**, 3135–3158 (2021). doi: 10.1007/s00382-021-05859-w
90. O. Hoegh-Guldberg *et al*., "Impacts of 1.5°C Global Warming on Natural and Human Systems. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty." V. Masson-Delmotte *et al*., Eds. (2018), pp. 175–311.
91. A. Erfanian, G. Wang, M. Yu, R. Anyah, Multimodel ensemble simulations of present and future climates over West Africa: Impacts of vegetation dynamics. *J. Adv. Model. Earth Syst.* **8**, 1411–1431 (2016). doi: 10.1002/2016MS000660
92. T. P. Hughes *et al*., Coral reefs in the Anthropocene. *Nature* **546**, 82–90 (2017). doi: 10.1038/nature22901; pmid: 28569801
93. K. Frieler *et al*., Limiting global warming to 2 °C is unlikely to save most coral reefs. *Nat. Clim. Chang.* **3**, 165–170 (2013). doi: 10.1038/nclimate1674
94. P. D. L. Ritchie, J. J. Clarke, P. M. Cox, C. Huntingford, Overshooting tipping point thresholds in a changing climate. *Nature* **592**, 517–523 (2021). doi: 10.1038/s41586-021-03263-2; pmid: 33883733
95. M. B. Osman *et al*., Globally resolved surface temperatures since the Last Glacial Maximum. *Nature* **599**, 239–244 (2021). doi: 10.1038/s41586-021-03984-4; pmid: 34759364
96. M. Meinshausen *et al*., Realization of Paris Agreement pledges may limit warming just below 2 °C. *Nature* **604**, 304–309 (2022). doi: 10.1038/s41586-022-04553-z; pmid: 35418633
97. T. Schneider, C. M. Kaul, K. G. Pressel, Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019). doi: 10.1038/s41561-019-0310-1
98. M. Rietkerk, R. Bastiaansen, S. Banerjee, J. Van De Koppel, Evasion of tipping in complex systems through spatial pattern formation. **374**, abj0359 (2021). doi: 10.1038/s41561-019-0310-1
99. T. M. Bury *et al*., Deep learning for early warning signals of tipping points. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106140118 (2021). doi: 10.1073/pnas.2106140118; pmid: 34544867
100. T. M. Lenton, Early warning of climate tipping points. *Nat. Clim. Chang.* **1**, 201–209 (2011). doi: 10.1038/nclimate1143

ACKNOWLEDGMENTS

Funding: This work was initiated and supported by the European Research Council Advanced Investigator project "Earth Resilience in the Anthropocene" ERC-2016-ADG-743080 (to D.I.A.M., A.S., S.C., I.F., and J.R.). This work is part of the Earth Commission which is hosted by Future Earth and is the science component of the Global Commons Alliance. The Global Commons Alliance is a sponsored project of Rockefeller Philanthropy Advisors, with support from Oak Foundation, MAVA, Porticus, Gordon and Betty Moore Foundation, Herlin Foundation, and the Global Environment Facility. (to D.I.A.M., J.F.A., R.W., B.S., S.L., and J.R.). This work was also supported by the Leverhulme Trust RPG-2018-046 (to T.M.L. and D.I.A.M.) and the Turing Fellowship (to T.M.L.). **Author contributions:** Conceptualization: D.I.A.M., A.S., I.F., S.E.C., and T.M.L. Methodology: D.I.A.M. and T.M.L. Investigation: D.I.A.M. Validation: A.S., J.F.A., R.W., B.S., I.F., and T.M.L. Visualization: J.F.A., B.S., and S.L. Funding acquisition: J.R. Data curation: D.I.A.M. Writing – original draft: D.I.A.M. and T.M.L. Writing – review and editing: D.I.A.M., A.S., J.F.A., R.W., B.S., S.L., I.F., S.E.C., J.R., and T.M.L. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7950
Materials and Methods
Supplementary Text
Tables S1 to S4
References (*101*–*312*)
Data S1

Submitted 21 December 2021; accepted 27 July 2022
10.1126/science.abn7950

RESEARCH ARTICLES

DOMESTICATION

The genomic history and global expansion of domestic donkeys

Evelyn T. Todd[1], Laure Tonasso-Calvière[1], Loreleï Chauvey[1], Stéphanie Schiavinato[1], Antoine Fages[1], Andaine Seguin-Orlando[1], Pierre Clavel[1], Naveed Khan[1,2], Lucía Pérez Pardal[3,4], Laura Patterson Rosa[5], Pablo Librado[1], Harald Ringbauer[6], Marta Verdugo[7], John Southon[8], Jean-Marc Aury[9], Aude Perdereau[9], Emmanuelle Vila[10], Matilde Marzullo[11], Ornella Prato[11], Umberto Tecchiati[11], Giovanna Bagnasco Gianni[11], Antonio Tagliacozzo[12], Vincenzo Tinè[13], Francesca Alhaique[12], João Luís Cardoso[14,15], Maria João Valente[16], Miguel Telles Antunes[17], Laurent Frantz[18,19], Beth Shapiro[20,21], Daniel G. Bradley[7], Nicolas Boulbes[22], Armelle Gardeisen[23], Liora Kolska Horwitz[24], Aliye Öztan[25], Benjamin S. Arbuckle[26], Vedat Onar[27], Benoît Clavel[28], Sébastien Lepetz[28], Ali Akbar Vahdati[29], Hossein Davoudi[30], Azadeh Mohaseb[28,30], Marjan Mashkour[28,30,31], Olivier Bouchez[32], Cécile Donnadieu[32], Patrick Wincker[9], Samantha A. Brooks[33], Albano Beja-Pereira[3,4,34,35], Dong-Dong Wu[36,37], Ludovic Orlando[1]*

Donkeys transformed human history as essential beasts of burden for long-distance movement, especially across semi-arid and upland environments. They remain insufficiently studied despite globally expanding and providing key support to low- to middle-income communities. To elucidate their domestication history, we constructed a comprehensive genome panel of 207 modern and 31 ancient donkeys, as well as 15 wild equids. We found a strong phylogeographic structure in modern donkeys that supports a single domestication in Africa ~5000 BCE, followed by further expansions in this continent and Eurasia and ultimately returning to Africa. We uncover a previously unknown genetic lineage in the Levant ~200 BCE, which contributed increasing ancestry toward Asia. Donkey management involved inbreeding and the production of giant bloodlines at a time when mules were essential to the Roman economy and military.

Domestic donkeys (*Equus asinus*) have facilitated the movement of goods and people for millennia, enabling trade and transport across a broad spectrum of landscapes (*1*). Despite their importance to ancient pastoral societies, little is known about the deep history of donkeys and the impact of human management on their genomes. This is most likely due to their undervalued status and loss of utility in modern industrialized societies. Donkeys are, however, extraordinary working animals and remain essential for developing communities, especially in semi-arid environments (*2*). Understanding their genetic makeup is not only key to assessing their contribution to human history but also to improving their local management in the future.

The current archaeological record of early donkeys is limited (*1*, *3*), which makes their domestic origins and spread through the world contentious. The reduced body size of zooarchaeological ass remains in Egypt at El Omari (4800 to 4500 BCE) and Maadi (4000 to 3500 BCE) has been interpreted as early evidence of domestication (*4*–*7*). Carvings on the Libyan palette that were found in Abydos, Egypt (3200 to 3000 BCE), depict lines of walking asses, cattle, and sheep, which also suggests a domestication context (*8*, *9*). Together with contemporary remains from the same region that show morphological evidence for load carrying (*10*), these findings suggest that donkeys could have been first domesticated within a range extending from the Northeastern Sahara, the Nile Valley, the Atbara River, and the Red Sea Hills to Eritrea. In this model, donkeys were domesticated by pastoralists to assist with mobility around 5500 to 4500 BCE because of the large-scale aridification of the Sahara (*1*). Independent evidence based on patterns of mitochondrial (*11*, *12*) and nuclear sequence variation (*13*) also points to the African origins of the donkey, owing to their closer proximity to African wild asses (*Equus africanus* spp.) than to Asian wild asses (*Equus hemionus* spp.).

However, candidate regions outside of Africa are also proposed as alternative domestication centers. In Ash Shuman (Yemen), for example, ass remains of disputed domestic status predate those from Egypt by 2000 years (~6500 BCE) (*14*). Likewise, textual, iconographic, and zooarchaeological material indicates a possible additional center in Mesopotamia during the fourth and third millennia BCE (*15*–*19*), a context in which first-generation hybrids of donkeys and Syrian onagers have been identified genetically (*20*). The segregation of mitochondrial variation in two main clades may also support a dual domestication process, for which the Nubian wild ass (*Equus africanus africanus*) is securely identified as the progenitor of Clade I (*11*, *12*). As for the ancestor of Clade II, it could either be the extinct Atlas wild ass (*Equus africanus atlanticus*), which was endemic to Northern Africa, or another undescribed subspecies that potentially ranged outside of Africa. Whether a single, maternally

[1]Centre d'Anthropobiologie et de Génomique de Toulouse (CAGT), CNRS UMR 5288, Université Paul Sabatier, Toulouse 31000, France. [2]Department of Biotechnology, Abdul Wali Khan University, Mardan 23200, Pakistan. [3]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal. [4]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal. [5]Department of Animal Science, Sul Ross State University, Alpine, TX 79830, USA. [6]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. [7]Smurfit Institute of Genetics, Trinity College Dublin, Dublin D02 PN40, Ireland. [8]Earth System Science Department, University of California, Irvine, CA 92697, USA. [9]Genoscope, Institut de biologie François Jacob, CEA, Université d'Evry, Université Paris-Saclay, Evry 91042, France. [10]Laboratoire Archéorient, Université Lyon 2, Lyon 69007, France. [11]Dipartimento di Beni Culturali e Ambientali, Università degli Studi di Milano, Milan 20122, Italy. [12]Bioarchaeology Service, Museo delle Civiltà, Rome 00144, Italy. [13]Soprintendenza archeologia belle arti e paesaggio per le province di Verona, Rovigo e Vicenza, Verona 37121, Italy. [14]ICArEHB, Campus de Gambelas, University of Algarve, Faro 8005-139, Portugal. [15]Universidade Aberta, Lisbon 1269-001, Portugal. [16]Faculdade de Ciências Humanas e Sociais, Centro de Estudos de Arqueologia, Artes e Ciências do Património, Universidade do Algarve, Faro 8000-117, Portugal. [17]Centre for Research on Science and Geological Engineering, Universidade NOVA de Lisboa, Lisbon 1099-085, Portugal. [18]Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich 80539, Germany. [19]School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4DQ, United Kingdom. [20]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA. [21]Howard Hughes Medical Institute, University of California, Santa Cruz, CA 95064, USA. [22]Institut de Paléontologie Humaine, Fondation Albert Ier, Paris / UMR 7194 HNHP, MNHN-CNRS-UPVD / EPCC Centre Européen de Recherche Préhistorique, Tautavel 66720, France. [23]Archéologie des Sociétés Méditéranéennes, Université Paul Valéry - Site Saint-Charles 2, Montpellier 34090, France. [24]National Natural History Collections, Edmond J. Safra Campus, Givat Ram, The Hebrew University, Jerusalem 9190401, Israel. [25]Archaeology Department, Ankara University, Ankara 06100, Turkey. [26]Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [27]Osteoarchaeology Practice and Research Center and Department of Anatomy, Faculty of Veterinary Medicine, Istanbul University-Cerrahpaşa, Istanbul 34320, Turkey. [28]Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements, Muséum National d'Histoire Naturelle, Paris 75005, France. [29]Provincial Office of the Iranian Center for Cultural Heritage, Handicrafts and Tourism Organisation, North Khorassan, Bojnord 9416745775, Iran. [30]Archaezoology section, Bioarchaeology Laboratory of the Central Laboratory, University of Tehran, Tehran CP1417634934, Iran. [31]Department of Osteology, National Museum of Iran, Tehran 1136918111, Iran. [32]GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, GET - Plateforme Génome & Transcriptome, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Castaneet-Tolosan Cedex 31326, France. [33]Department of Animal Science, UF Genetics Institute, University of Florida, Gainesville, FL 32610, USA. [34]DGAOT, Faculty of Sciences, Universidade do Porto, Porto 4169-007, Portugal. [35]Sustainable Agrifood Production Research Centre (GreenUPorto), Universidade do Porto, Vairão 4485-646, Portugal. [36]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650201, China. [37]Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China.
*Corresponding author. Email: ludovic.orlando@univ-tlse3.fr

inherited marker captures the whole complexity of underlying ancestries can, however, be questioned after recent results from other animals [e.g., horses (*21*)]. Furthermore, previous analyses of nuclear genetic variation in African and non-African donkeys have failed to disentangle their origins (*13*, *22*). Overall, this lack of consensus between genetic and archaeological data means that the geographic and temporal origin of donkeys and whether they were domesticated more than once remain uncertain. The global spread of donkeys is also unclear, as their worldwide patterns of genomic diversity lack extensive characterization.

Modern donkeys originated in Africa and spread into Eurasia

To address these issues, we sequenced 49 modern donkey genomes from underrepresented regions and combined these with 158 publicly available genomes to capture worldwide diversity (Fig. 1A and table S1) (*13*, *23*–*25*). We constructed a fine-scale recombination map from genomes that encompassed all phylogenetic groups, which we used to phase 13,013,551 variants (fig. S1 and tables S3 and S4). Principal component analysis (*26*) revealed strong geographical substructuring, with donkeys from Africa, Europe, and Asia forming distinct genetic clusters (Fig. 1B and figs. S2 and S3).

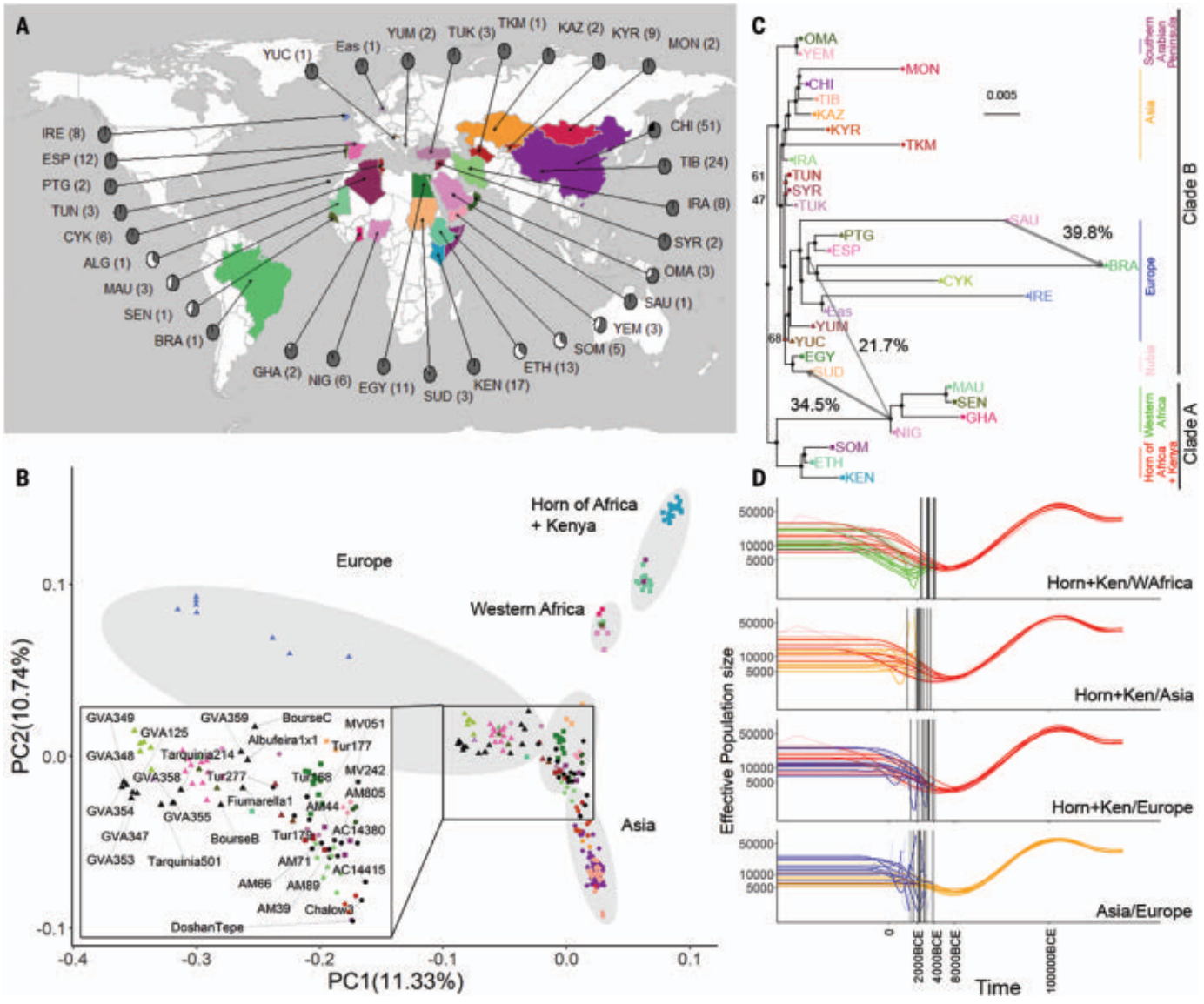


Fig. 1. Modern donkey dataset and population evolutionary history. (**A**) Number and geographical distribution of modern donkey samples (*n* = 207). Pie charts show the ADMIXTURE proportion of domestic ancestry (gray), African wild ass ancestry (white), and kiang ancestry (black) averaged across all individuals from each country (*56*). For visualization, the total surface of each pie chart is scaled to 2%. (**B**) Smartpca (*57*) of modern donkeys, with the imputed ancient samples in black. PC, principal component. (**C**) Treemix phylogeny of modern domesticates (excluding individuals with high wild introgression, *n* = 201) (*27*). Node supports are estimated from 100 bootstrap pseudo-replicates (confidence <90% in red). The percentage values indicate admixture proportions inferred from Treemix (*27*). (**D**) SMC++ demographic trajectories (colored) and split time estimates (black) for pairs of main geographic regions (*28*), repeating the analysis on two datasets of three individuals per population (the second dataset is shown in semitransparency). Modern donkeys are colored and shaped according to geographical location and continents in all panels. Country Codes: Algeria (ALG), Brazil (BRA), China (CHI), Canary Islands (CYK), Denmark (Eas), Egypt (EGY), Spain (ESP), Ethiopia (ETH), Ghana (GHA), Iran (IRA), Ireland (IRE), Kazakhstan (KAZ), Kenya (KEN), Kyrgyzstan (KYR), Mauritania (MAU), Mongolia (MON), Nigeria (NIG), Oman (OMA), Portugal (PTG), Saudi Arabia (SAU), Senegal (SEN), Somalia (SOM), Sudan (SUD), Syria (SYR), Tibet (TIB), Turkmenistan (TKM), Turkey (TUK), Tunisia (TUN), Yemen (YEM), Croatia (YUC), Macedonia (YUM).

A Treemix phylogenetic reconstruction that grouped modern donkeys according to sampling locations (*27*) confirms the earliest split between African (Clade A) and mostly non-African donkeys (Clade B) (Fig. 1C). Further structure within Clade A separates donkeys from the Horn of Africa (Ethiopia and Somalia) plus Kenya from those from Western Africa (Ghana, Mauritania, Nigeria, and Senegal). Within Clade B, we find another major divergence between European and Asian donkeys, with east-to-west affinities in Europe from the Balkans (Croatia and Macedonia) to Iberia, Denmark, and Ireland. Conversely, Asian subpopulations show west-to-east substructuring from Iran and Central Asia to China and Mongolia. Combined, these findings suggest expansions from a central source into both continents.

In Clade B, some of the most basal donkeys are from the Southern Arabian Peninsula (Oman and Yemen), whereas the single donkey from Saudi Arabia analyzed here shows European affinities, which is indicative of a secondary translocation. Similarly, the Pega donkey from Brazil is nested within Iberia, which mirrors the colonization history of the Americas. Clade B also includes donkeys from Nubia (Egypt and Sudan) that show affinities to the Levant (Syria) and Anatolia (Turkey), as well as donkeys from Maghreb (Tunisia), with closer genetic proximity to European subpopulations. This suggests gene flow into Africa from donkeys that were native to Anatolia and the Levant, but not to the Arabian Peninsula. Overall, this phylogenetic reconstruction is compatible with both models of donkey domestication: a unique origin in Africa followed by dispersals out and back, or dual origins in Africa and the Southern Arabian Peninsula.

The unique origin model posits a demographic expansion in Africa first and subsequent waves into Europe and Asia. By contrast, dual origins would result in an earlier split of demographic trajectories between African and Eurasian subpopulations, given their deep phylogenetic divergence. To test these, we first performed demographic modeling using SMC++ (*28*), which revealed the first expansion around 5200 BCE (7186 ± 742 years ago), in line with archaeological evidence of domestication occurring at this time (Fig. 1D and fig. S6). Additionally, when modeled from a possible African source, SMC++ trajectories indicated more recent and nearly coincidental expansions into Asia around 2600 BCE (4573 ± 577 years ago) and Europe around 2800 BCE (4806 ± 671 years ago) (Fig. 1D). This is in line with the unique origin model and the earliest archaeological evidence of donkeys in Asia (Iranian Plateau and the Indus Valley) and Europe (Portugal, Greece, and Cyprus) in the mid-to-late third millennia BCE (*29*–*34*).

Furthermore, the Yemen and Oman subpopulations do not branch basal to Clade B according to fineSTRUCTURE (*35*), in contrast to the expectations of the dual origins model, but within Asian subpopulations (Fig. 2, A and B). Lastly, the pairwise genetic distances between Ethiopia and non-African subpopulations were greater than those from Yemen (fig. S5). They both increased linearly with geographic distances and supported identical dispersal rates (fig. S5; $p = 0.775$), which is in line with a single wave of expansion at a constant pace. Therefore, our analyses support an early domestication in Africa that spread at an even rate into the Arabian Peninsula and Eurasia and flowed back into Nubia and Maghreb. Modern subpopulations from the Horn of Africa and Kenya so far best represent the descendants of earliest donkeys.

Ancient donkey genomes reveal early and rapid dispersal into Asia and secondary contacts between Europe and Western Africa

The patterns of genetic variation within modern subpopulations may reflect recent breeding history rather than early domestication (*36*). Additionally, they could underrepresent the contribution of lineages that were once important but have since declined (*37*). Dating population splits also assumes constant yet unknown generation intervals. To address these caveats and validate the domestication history reconstructed above, we generated a genomic time series spanning the past ~4000 years that included 31 ancient donkeys from 11 different sites, which ranged from the Atlantic shores (Portugal) to Central Asia (Eastern Iran/Turkmenistan) (Fig. 3A and table S2).

Ancient genomes sequenced to 0.77- to 5.05-fold coverage (table S6) were analyzed by using two complementary methods: pseudo-haploidization following (*21*) that resulted in 4,833,570 nucleotide transversion sites and genotype imputation following (*38*), at 7,161,029 polymorphic sites present at >5% frequency in modern donkeys (*38*, *39*). Imputation accuracy was confirmed by the high consistency rates between imputed and observed genotypes after down-sampling of high-coverage modern genomes and downstream analyses that were largely consistent with those that were based on pseudo-haploidized data (figs. S7 to S15 and table S6).

The three oldest samples from our dataset consist of donkeys from Anatolia (Acemhöyük, Turkey), which were radiocarbon dated to 2564 to 2039 BCE. Their age and phylogenetic placement within Clade B confirm an early expansion out of Africa by ~2500 BCE, which is in agreement with SMC++ time estimates (Fig. 2A and fig. S10). These samples, as well as a donkey from Eastern Iran/Turkmenistan that is affiliated to the Bactria–Margiana Archaeological Complex (BMAC) (~2050 BCE; sample Chalow3), branch before the formation of modern subpopulations from Central Asia (Kazakhstan, Kyrgyzstan, Turkmenistan) and Eastern Asia (China, Mongolia, Tibet) (Fig. 2C). These subpopulations thus diverged after 2050 BCE, but potentially before the radiocarbon age of the donkey from Doshan Tepe (1049 to 928 BCE), which appears closer to modern subpopulations from Central Asia in one Treemix analysis (Fig. 2D and fig. S10).

Ancient samples from Iran (Shahr-i-Qumis, 800 BCE to 800 CE), including one Sassanid (sample AM805), are not more closely related to Central than to Eastern Asian modern subpopulations, although their exact phylogenetic placement remains unclear (Fig. 2, A and B, and fig. S10). Their fineSTRUCTURE affinities to modern Iran, Anatolia (Turkey), the Levant (Syria), and Maghreb (Tunisia) support different genetic ancestry profiles from those inferred at the nearby site of Doshan Tepe. This indicates a population turnover in Iran after ~1000 BCE but before ~500 CE, corresponding to the radiocarbon time interval of Doshan Tepe and a single specimen from Shahr-i-Qumis.

Notably, all our ancient specimens from Europe cluster within modern European domesticates, supporting differentiation within this continent before the oldest European samples analyzed (Tarquinia, 803 to 412 BCE, ~2500 years ago; Fig. 3C). However, a donkey from a Roman context in Marseille, a major seaport trading center in Southern France (Centre Bourse Marseille, 0 to 500 CE), displayed strong genetic affinities with modern individuals from Western Africa (Fig. 3, B and D). Additionally, SNP and haplotype sharedness with modern Western Africa were also found in European donkeys from the Islamic era in Portugal (Albufeira, 1228 to 1280 CE) and Roman times in Northern France (Boinville-en-Woëvre, 200 to 500 CE) (Fig. 3, B and E). This reveals multiple contacts between Europe and Western Africa from the Antiquity to Middle Ages. Despite ancient European donkeys showing Western African ancestry, these contacts have affected Western Africa more than Europe, in line with Treemix inferring gene flow predominantly in this direction rather than the reverse (Fig. 1C). All modern Irish donkeys and the two Etruscan samples from Tarquinia are devoid of Western African ancestry. This suggests the preservation of old European genetic lineages, at least in some modern subpopulations of this continent.

Donkey management involved inbreeding and introgression from divergent lineages

Inbreeding is a common reproductive strategy for breeding animals with desirable traits (*40*).

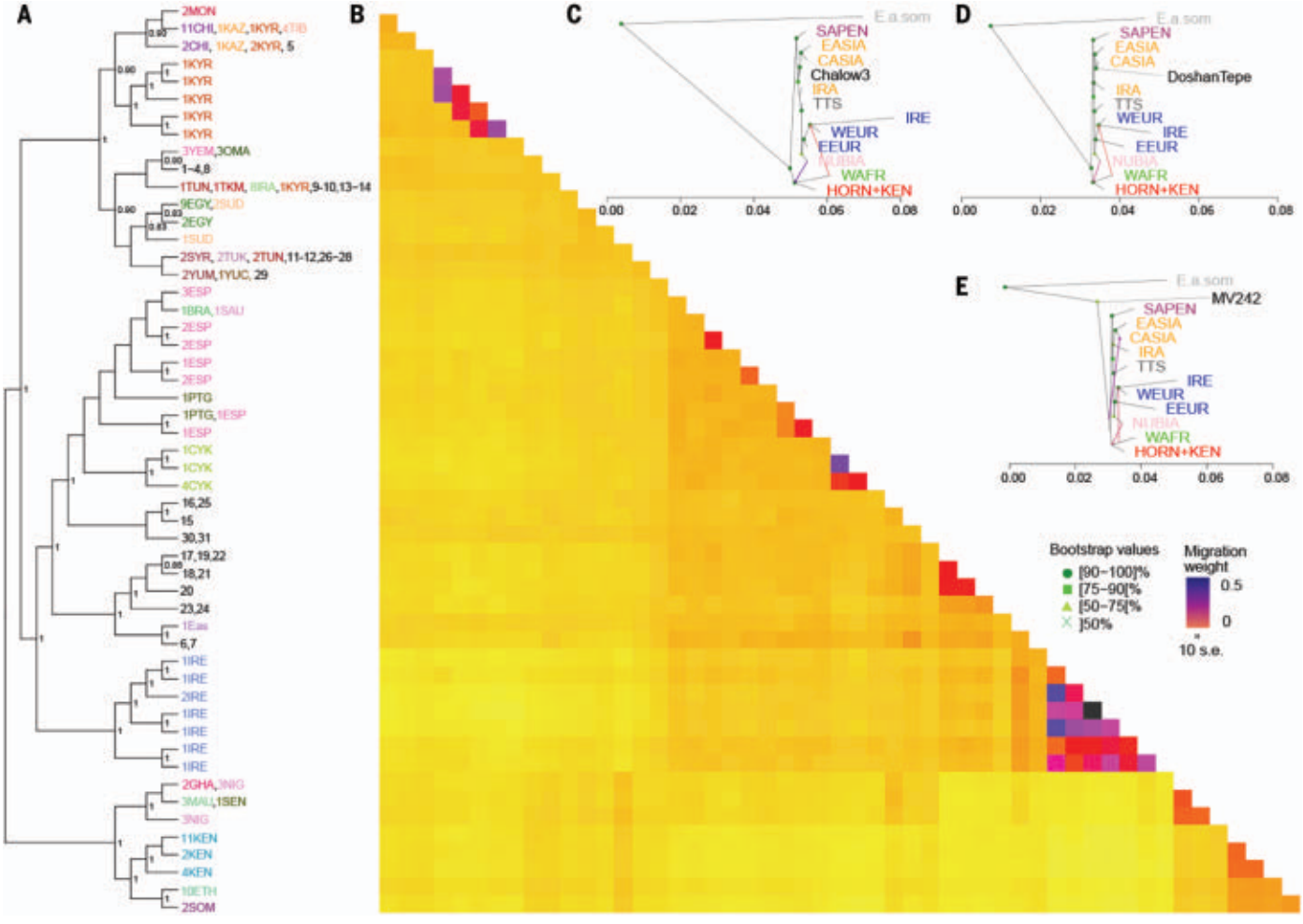


Fig. 2. Haplotype sharedness and phylogenetic placement of ancient European donkeys. (**A**) Haplotype sharedness clustering of modern (n = 168) and ancient donkeys (n = 31) reconstructed using fineSTRUCTURE (*35*). Modern domesticates are colored following Fig. 1, and ancient individuals are numbered according to Fig. 3A. Cluster supports are shown in as probabilities on each node if >0.8. MV242 placement is incongruent with Treemix (Fig. 2E) because of the limited representation of divergent ancestries in the modern reference panel used for imputation. (**B**) Co-ancestry matrix based on haplotype sharedness. Co-ancestry values are averaged for co-clustered individuals. (**C** to **E**) Treemix phylogenies of three ancient specimens shown in black (C: Chalow3, D: Doshan Tepe, and E: MV242) placed within the subpopulations defined in Fig. 1C (*27*). Branches that are not scaled are shown as dashed lines. s.e., standard error.

To assess whether ancient donkey breeders made use of inbreeding, we measured the proportion of autosomal runs of homozygosity (ROH) using three independent techniques, all of which provided consistent results (figs. S13 and S14). We detected inbreeding, but no significant changes in levels between modern and ancient donkeys (Wilcoxon rank sum test, p = 0.395) (Fig. 4, A, B, and C). Conversely, modern horses show higher inbreeding levels than their ancient counterparts (Wilcoxon rank sum test, $p < 0.001$), which mirrors previous reports of reduced heterozygosity and increased deleterious mutation load in recent times (Fig. 4, D, E, and F, table S11) (*21*, *41*). Longer ROH tracts are more common in modern horses and donkeys than in the past, which is consistent with inbreeding from closer generations in their genealogies (Fig. 4, C and F). Overall, our analyses support recent major changes in reproductive management inflating inbreeding in horses, but not in donkeys.

Admixture modeling suggests ongoing introgression from African wild asses into modern donkeys from Africa and the Southern Arabian Peninsula (with between 0.24 to 6.99% of admixture) (Fig. 1A, fig. S4, and table S5). This is in line with free-ranging local management practices that allow for continued interbreeding with wild and feral subpopulations (*4*, *42*). The limited but substantial amount of wild genetic material from kiangs in one modern donkey from China also supports admixture between taxa generally regarded as separate species. This confirms previous reports of mitochondrial introgression (*43*) and genomic admixture despite different karyotypes (*24*). Notably, all but one ancient donkey (sample Tur168) carried remnants of outgroup material (0.21 to 4.15%; Fig. 3B), which potentially results from recent range contractions of wild subpopulations and ancient management practices providing more opportunities for wild introgression.

The genome of MV242, a donkey from Israel dating to the Hellenistic period (350 to 58 BCE), displayed the largest fraction of divergent genetic material (Fig. 3B, 4.15% ± 0.019). In Treemix, this sample showed a deeper placement than all donkeys present in our panel, except the Somali wild ass [*Equus africanus somaliensis* (*E.a.som*)] (Fig. 2E). Significantly positive f4 (*E.a.som*, MV242; Horn+Ken, x) statistics revealed MV242-related genetic ancestry in some modern subpopulations (x), especially toward Central and Eastern Asia (Fig. 5E). This ancestry was already present in the BMAC sample from Iran (sample Chalow3, Fig. 5F), which indicates contact ~2050 BCE at the

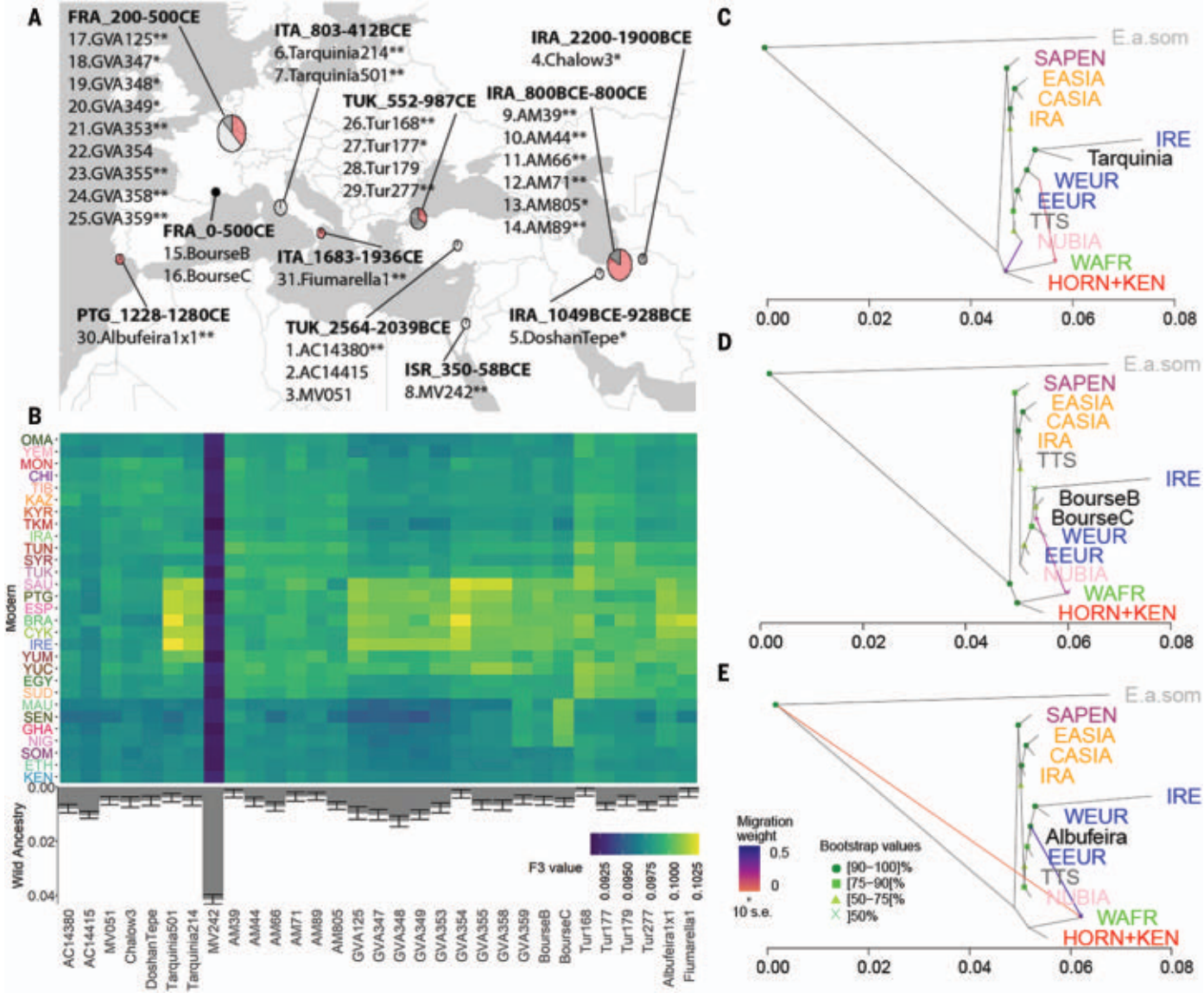


Fig. 3. Ancient donkey dataset, genetic affinities to outgroups, and modern donkeys. (**A**) Geographical distribution, estimated age, and sample names of ancient donkeys (n = 31). Pie charts represent the proportion of individuals with dun coat color (white), heterozygotes (gray), and derived coat color (red) at each site. Genotype probabilities ≥0.99 are denoted with ** and ≥0.9 with *. (**B**) Heatmap displaying outgroup f3-statistics in the form of (modern, ancient; kiang) (*58*). Bar charts represent the proportion of wild ancestry (kiang, onager, zebra, *E.a.som*) in each ancient individual with standard errors estimated from ADMIXTURE with 100 bootstrap pseudo-replicates (*56*). (**C** to **E**) Treemix phylogenies of ancient specimens from three archaeological sites shown in black [C: Tarquinia (Tarquinia214, Tarquinia501), D: Bourse (BourseB, BourseC), E: Albufeira 1x1] placed within the subpopulations defined in Fig. 1C (*27*). Branches that are not scaled are shown as dashed lines. s.e., standard error.

latest. It was, however, absent in Acemhöyük at that time, which suggests that the MV242 divergent lineage ranged into Eastern Iran/Turkmenistan, but not Turkey. This lineage also left genetic material in modern Anatolia, the Levant, Nubia, and Maghreb, but not in Western Africa, which is consistent with donkeys carrying MV242-related ancestry flowing back into some African regions. In addition, this ancestry was also present in Southwestern European subpopulations (Canary Islands, Portugal, and Spain), but neither in the modern Balkans and Ireland nor in any ancient European sample analyzed here (Fig. 5, E and F). Combined, our results suggest a range for the MV242-related lineage from the Levant into Asia rather than Europe and Africa.

Despite its divergent genetic makeup, MV242 carries a mitochondrial haplotype characteristic of Clade II (Fig. 5A). Our tip-calibrated coalescent analyses revealed that the time to the most common recent ancestor of that Clade was 32,226 BCE and not 332,580 to 142,980 BCE (Fig. 5B), as previously reported (*12*, *44*). Because the same holds true for Clade I, both clades could have coexisted in sympatry 25,000 years later when donkeys were first domesticated (Fig. 5B). Additionally, no phylogeographic structure is apparent in patterns of mitochondrial variation, both in modern and ancient subpopulations, as ancient specimens from Asia and Europe, sometimes from the same archaeological sites, were placed across both clades (Fig. 5A). Y-chromosomal variation was also associated with little, if any, population structure (Fig. 5, C and D). Combined, our results dismiss mitochondrial DNA and the Y-chromosome as reliable markers of domestication history in donkeys.

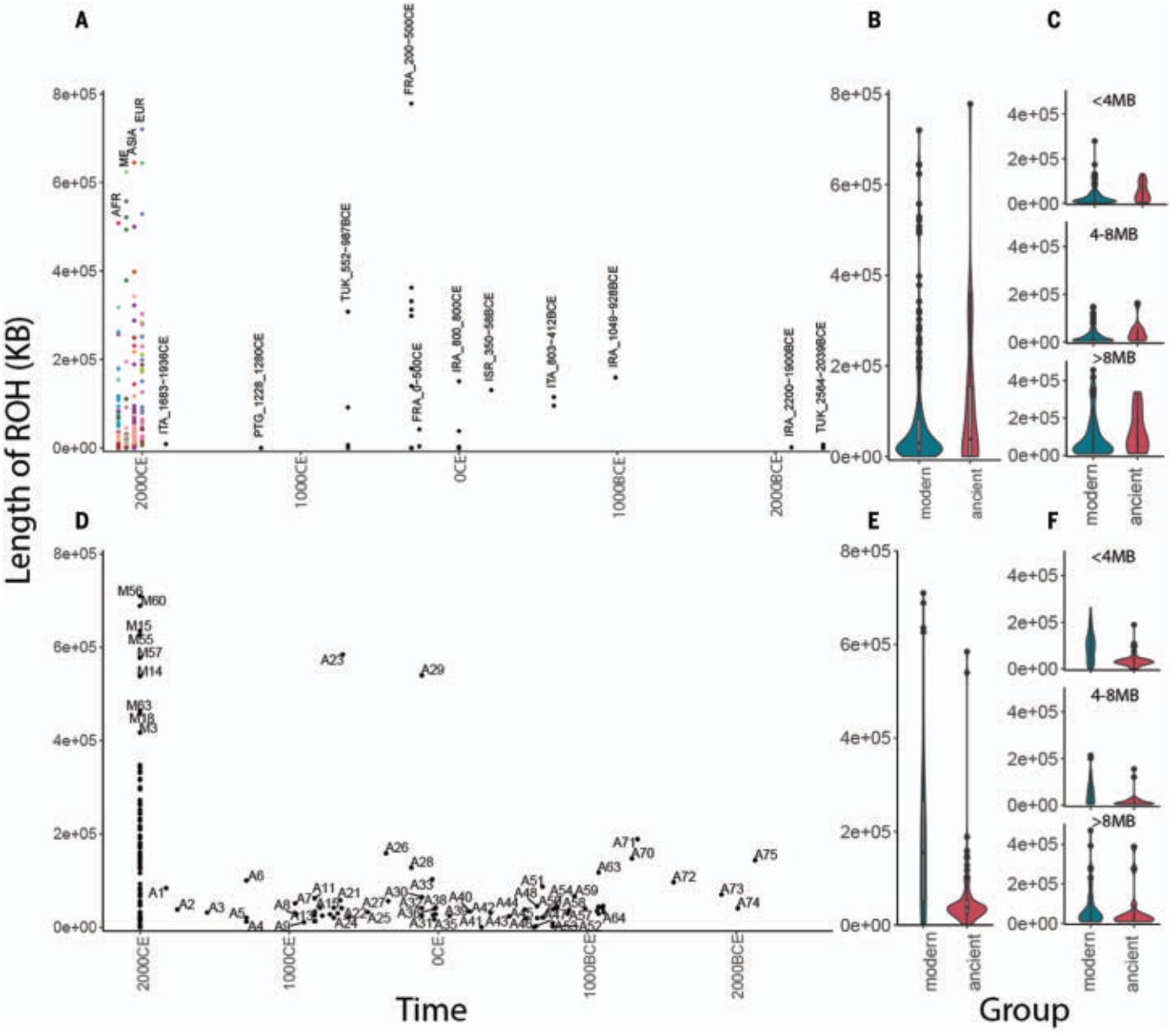


Fig. 4. Inbreeding in domestic donkeys and horses. (**A**) Distribution of total ROH length in modern versus ancient donkeys. (**B**) Total length of ROH in donkey genomes through time. (**C** and **D**) Same as (A) and (B), but for 79 modern and 75 ancient horses. ROH tracts were identified using the program ngsF-HMM (*59*).

Romans bred improved donkeys for producing mules essential for their military power and economy

Beyond documenting domestication history at the global scale, our genomic dataset also included three jennies (females) and six jacks (males) from the same archaeological site (Boinville-en-Woëvre) (Fig. 3A). These were found in a dedicated farming area of a Roman villa, providing insights into local management practices in Roman Northern France (200 to 500 CE). One jack (sample GVA349) appeared particularly inbred, with long ROH indicative of recent consanguinity (Fig. 4A), and was genetically related to four jacks and one jenny (family group GVA1, including samples GVA125, GVA347, GVA348, GVA349, GVA353, and GVA354; table S10). Additionally, two jennies showed genetic relatedness coefficients equivalent to full siblings (family group GVA2, including samples GVA355 and GVA358; table S10). This indicates breeding management within close kin, potentially aimed at selecting for desirable traits. Genotype imputation at *TBX3* (*13*) revealed the presence of dun and derived colored coats, but no evidence for the dominant alleles associated with white spots or long hair was found in the sequence alignments at *KIT* (*45*) and *FGF5* (fig. S11 and tables S7 to S9) (*46*). The latter two phenotypes are, however, relatively common in modern breeds from France, which suggests post-Roman selection for these traits.

The abundance of donkeys at Boinville-en-Woëvre stands as an exception in Roman France, as mules dominated all other assemblages from this time (*47*). Contemporaneous Roman sites report mules of a large and uniform size, indicating selective breeding in the parental species for expensive animals of exceptional stature [Varro (2, 6)] (*48*). Morphometric measurements previously revealed five donkeys from family group GVA1 as giant (148 to 156 cm at the withers) (*47*). We found that sample GVA359 had a similarly large size

(144 cm) and genetic affinities to Western Africa. This may indicate restocking to enhance body size from distant bloodlines carrying divergent ancestry or from wild populations.

Outgroup admixture was significantly higher at Boinville-en-Woëvre than in other ancient donkeys except the divergent MV242 specimen (p = 0.045). Significantly negative f4(kiang, MV242; Fiumarella1, Boinville-en-Woëvre) statistics support restocking into family group GVA1 only, from a lineage more divergent than MV242 (Fig. 5G). Additionally, f4(kiang, *E.a.som*; Fiumarella1, GVA1) statistics reject unbalanced allele sharedness between *E.a.som* and GVA1, ruling out restocking from *E.a.som* or more divergent populations (Fig. 5H). Combined, these findings uncover a lineage, phylogenetically intermediate between MV242 and *E.a.som*, that contributed to the genetic makeup of some Roman donkeys at Boinville-en-Woëvre. Together with the evidence of genetic relatedness and inbreeding, this suggests Boinville-en-Woëvre as a likely mule production center that maintained the bloodlines of giant donkeys

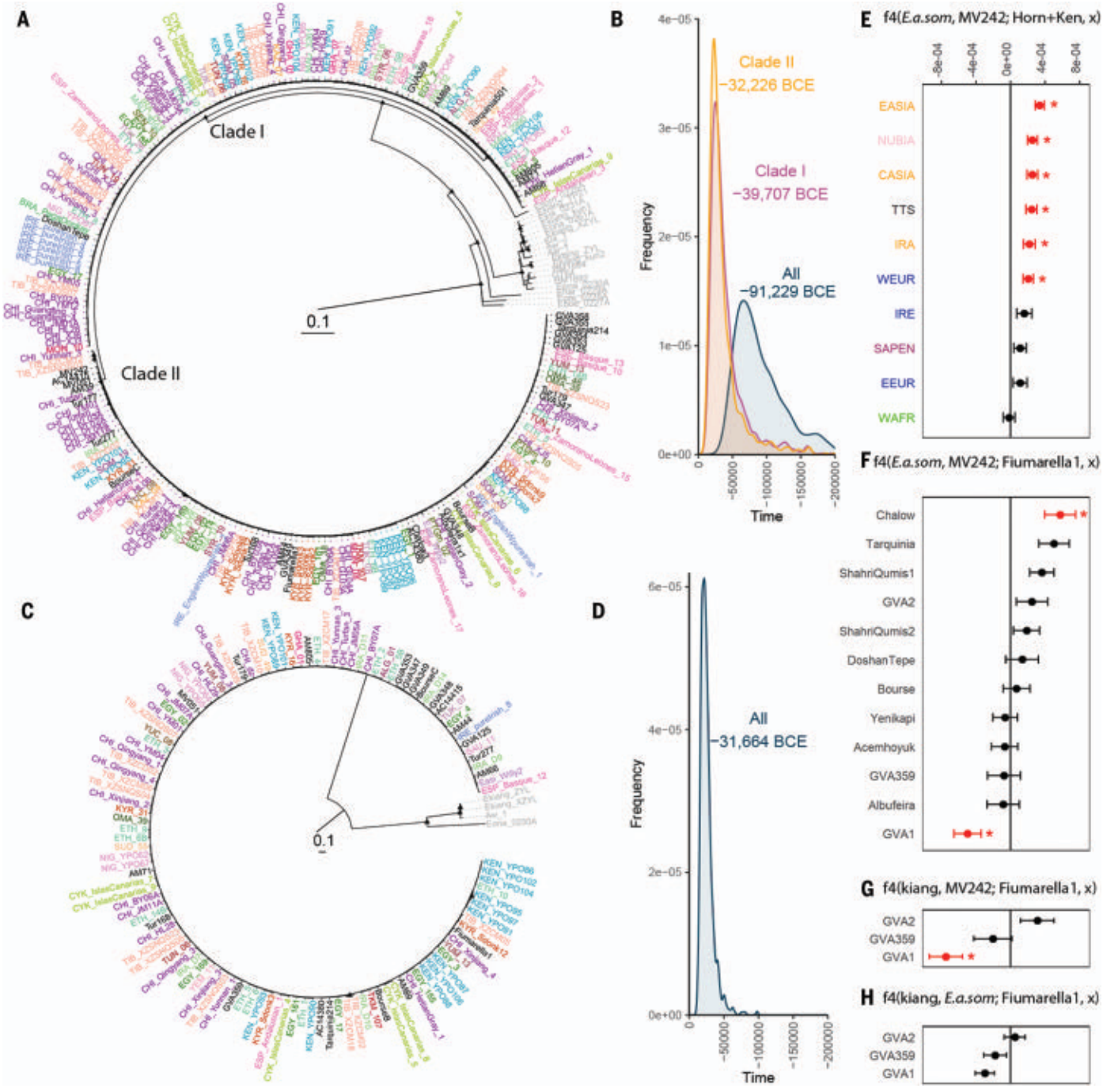


Fig. 5. Uniparental marker phylogenies and introgression of divergent lineages. (**A**) Mitochondrial phylogeny constructed using IQ-TREE (*60*) with 100 bootstrap pseudo-replicates marked with a black triangle if >90%. (**B**) Posterior distributions of the time to the most recent common ancestors of all mitochondrial haplotypes, Clade I and Clade II labeled with their modes. (**C** and **D**) Same as (A) and (B) for the Y-chromosome. (**E** to **H**) f4-statistics (*58*) exploring the genetic contribution of divergent lineages into modern and ancient donkeys. Z scores were corrected for multiple testing, and red bars with asterisks show p < 0.05.

that were selected through familial breeding and restocking. This center may illustrate how Romans sustained the enormous demand for mules, which is documented in the nearby Rhine frontier (*49*), and fueled transportation networks throughout the Empire (*47*).

Discussion

Our study solves long-standing debates about donkey domestication. We support domestication starting from a unique African source ~5000 BCE. Donkeys subsequently spread into Eurasia from ~2500 BCE, and Central and Eastern Asian subpopulations differentiated ~2000 to 1000 BCE. Genetic affinities characteristic of modern Western Europe were already formed by 500 BCE. After early domestication, African donkeys further differentiated in the West and the Horn of Africa plus Kenya but also received streams of genetic ancestry from Western Europe as well as a region encompassing the Levant, Anatolia, and Mesopotamia. Donkey domestication involved limited but detectable wild introgression. It did not entail inflated inbreeding in recent times, in contrast to horses. The processes of donkey and horse domestication differed substantially, as horses were domesticated twice (*50*) and rapidly spread across Eurasia from the lower Don-Volga region ~2000 BCE (*21*). Their regional differentiation remained relatively limited owing to strong connectivity at continental distances early on and until oriental bloodlines were propagated throughout the world during the past 1000 years (*41*, *51*). The extent to which the different domestication trajectories of donkeys and horses were only driven by their respective roles in human societies or also reflected management practices adapted to their respective mating and social behavior (*52*) remains to be explored.

This work clarifies global patterns of donkey domestication and movements but also highlights many directions for future research. For example, it remains unknown whether domestic donkeys only dispersed out of Africa by land through the Sinai Peninsula or across the Red Sea from Ethiopia to Yemen. Additionally, modern subpopulations from the Horn of Africa plus Kenya were found to be the first expanding. This may suggest early domestication there or that donkeys domesticated elsewhere in Africa entering the region more recently. Further research is needed to clarify the timing of pastoral spread into the Red Sea Sudanese region and the Horn of Africa. Current dates range from ~2500 BCE in Ethiopia and Eritrea (*53*) to ~3000 BCE in Northern Kenya (*54*). Donkeys are not present in the archaeological record of Western Africa before the beginning of the common era, either (*55*), which postdates by 3000 years the time when donkey populations from the Horn of Africa plus Kenya and Western Africa are inferred to have split genetically. This may indicate an early yet undocumented arrival in the region or a slow migration westward, only reaching the modern range later. Improving the current African archaeological record thus appears paramount to refining the exact context underlying early donkey domestication and subsequent population movements.

Further genomic studies in other regions would also largely benefit the understanding of donkey diversity and history. Resolving the genetic structure of equine remains from the third millennium BCE of Southwest Asia will be challenging because of postmortem DNA decay but essential to mapping the geographic range of the divergent lineage identified here (sample MV242), as well as to understanding dispersal mechanisms in greater detail. The same holds true for Chalcolithic and Bronze Age Europe, which remain genetically undocumented in our dataset, and onwards. Developing genetic knowledge of ancient European donkeys will further clarify patterns of exchange across the Mediterranean region, including during and after Roman times, as revealed in this study. It will also provide insights into the dispersal mechanisms underpinning the genetically supported presence of donkey remains in Portugal ~2200 BCE (*33*). Genetic characterization of local archaeological sites at the population scale may uncover additional mule breeding centers other than the one reported here. This will shed light on the diversity of breeding management strategies that were developed by Romans to supply their continental-wide economy and military with adequate animal resources (*49*). For now, both the absence of mules and the rarity of horse mares at Boinville-en-Woëvre (*47*) suggest that mares were brought in for mating before returning pregnant to their owners. Alternatively, donkey breeders may have visited other farms with their jacks to cover mares.

Efforts should continue to characterize the modern donkey diversity around the world, especially in Saudi Arabia, which is currently characterized by a single individual, as well as in Africa, for which no populations located south of the Equator have been sampled. Such efforts may not only refine the historical legacy of past populations into the modern world but also uncover the genetic basis of desert adaptations, which could prove invaluable for future donkey breeding in the face of global warming.

REFERENCES AND NOTES

1. P. Mitchell, *The Donkey in Human History: An Archaeological Perspective* (Oxford Univ. Press, 2018).
2. S. L. Norris, H. A. Little, J. Ryding, Z. Raw, *PLOS ONE* **16**, e0247830 (2021).
3. B. Kimura, F. Marshall, A. Beja-Pereira, C. Mulligan, *Afr. Archaeol. Rev.* **30**, 83–95 (2013).
4. F. Marshall, *Rethinking Agriculture: Archeological and Ethnoarcheological Perspectives* (Left Coast Press, 2007).
5. S. Bokonyi, in *Studi di paletnologia in onore di Salvatore M. Puglisi*, M. Liverani, A. Palmieri, R. Peroni, Eds. (Università di Roma La Sapienza, 1985), pp. 495–499.
6. J. Boesneck, A. Von den Driesch, in *El Omari: A Neolithic Settlement and Other Sites in the Vicinity of Wadi Hof, Helwan*, F. Debono, B. Mortensen, Eds. (Philipp von Zabern, 1990), pp. 99–107.
7. J. Boesneck, A. von den Driesch, R. Ziegler, in *Maadi III: The Non-lithic Small Finds and the Structural Remains of the Predynastic Settlement*, I. Rizkana, J. Seeber, Eds. (Philipp von Zabern, 1989), pp. 87–128.
8. K. M. Cialowicz, *Les palettes Egyptiennes aux motifs zoomorphes et sans decoration. Etudes de l'art predynastique* (Jagiellonian University, 1991).
9. D. J. Brewer, D. B. Redford, S. Redford, *Domestic Plants and Animals: The Egyptian Origins* (Aries and Phillips, 1994).
10. S. Rossel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3715–3720 (2008).
11. A. Beja-Pereira *et al.*, *Science* **304**, 1781 (2004).
12. B. Kimura *et al.*, *Proc. Biol. Sci.* **278**, 50–57 (2011).
13. C. Wang *et al.*, *Nat. Commun.* **11**, 6014 (2020).
14. M. Cattani, S. Bokonyi, in *Essays on the Late Prehistory of the Arabian Peninsula. Serie Orientale Romana XCIII*, S. Cleuziou, T. Maurizio, Z. Juris, Eds. (Istituto Italiano per l'Africa e l'Oriente, 2002).
15. R. H. Meadow, H. P. Uerpmann, *Equids in the Ancient World, vol. 2* (Dr Ludwig Reichert Verlag, 1991).
16. J. Clutton-Brock, *Mammal Rev.* **22**, 79–85 (1992).
17. J. Zarins, R. Hauser, The domestication of equidae in third-millennium BCE Mesopotamia. *Cornell University Studies in Assyriology and Sumerology* **24**, XI + 432 (2014).
18. E. Vila, in *Equids in Time and Space: Papers in Honour of Véra Eisenmann*, M. Mashkour, Ed. (Oxbow, 2006), pp. 101–123.
19. J. Boessneck, A. von den Driesch, U. Steger, *Baghdader Mitteilingen* **15**, 149–189 (1984).
20. E. A. Bennett *et al.*, *Sci. Adv.* **8**, eabm0218 (2022).
21. P. Librado *et al.*, *Nature* **598**, 634–640 (2021).
22. S. Rosenbom *et al.*, *Mol. Phylogenet. Evol.* **85**, 88–96 (2015).
23. G. Renaud *et al.*, *Sci. Adv.* **4**, eaaq0392 (2018).
24. H. Jónsson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18655–18660 (2014).
25. L. Zeng *et al.*, *Zool. Res.* **42**, 450–460 (2021).
26. A. L. Price *et al.*, *Nat. Genet.* **38**, 904–909 (2006).
27. J. K. Pickrell, J. K. Pritchard, *PLOS Genet.* **8**, e1002967 (2012).
28. J. Terhorst, J. A. Kamm, Y. S. Song, *Nat. Genet.* **49**, 303–309 (2017).
29. D. Reese, *Mediterr. Archaeol. Archaeom. Int. J.* **13**, 289–320 (2013).
30. B. A. Knapp, *Biblic. Archaeol.* **55**, 52–72 (1992).
31. S. Ratnagar, *Trading Encounters: from the Euphrates to the Indus in the Bronze Age* (Oxford Univ. Press, 2004).
32. M. A. Zeder, in *Equids in the Ancient World*, R. H. Meadow, H. P. Uerpmann, Eds. (Reichert, 1986), pp. 164–193.
33. J. L. Cardoso, J. T. Vilstrup, V. Eisenmann, L. Orlando, *J. Archaeol. Sci.* **40**, 4483–4490 (2013).
34. S. Amiri, M. Mashkour, F. Mohaseb, R. Naseri, in *Archaeozoology of Southwest Asia and Adjacent Areas XIII: Proceedings of the Thirteenth International Symposium, University of Cyprus, Nicosia, Cyprus, June 7-10, 2017*, J. Daujat, A. Hadjikoumis, R. Berthon, J. Chahoud, V. Kassianidou, J. D. Vigne, Eds. (2021), pp. 199–219.
35. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, *PLOS Genet.* **8**, e1002453 (2012).
36. L. Girdland Flink *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6184–6189 (2014).
37. L. A. F. Frantz, D. G. Bradley, G. Larson, L. Orlando, *Nat. Rev. Genet.* **21**, 449–460 (2020).
38. R. Hui, E. D'Atanasio, L. M. Cassidy, C. L. Scheib, T. Kivisild, *Sci. Rep.* **10**, 18542 (2020).
39. B. L. Browning, S. R. Browning, *Am. J. Hum. Genet.* **98**, 116–126 (2016).
40. T. N. Kristensen, A. C. Sørensen, *Anim. Sci.* **80**, 121–133 (2007).
41. A. Fages *et al.*, *Cell* **177**, 1419–1435.e31 (2019).
42. F. Marshall, L. Weissbrod, in *Tracking Down the Past. Ethnohistory Meets Archaeozoology*, G. Grupe, J. Peters, G. McGlynn, Eds. (M. Leidor, 2009), pp. 59–79.
43. E. A. Bennett *et al.*, *PLOS ONE* **12**, e0174216 (2017).
44. L. Wang *et al.*, *Front. Genet.* **12**, 759831 (2021).
45. B. Haase, S. Rieder, T. Leeb, *Anim. Genet.* **46**, 321–324 (2015).
46. R. Legrand, L. Tiret, M. Abitbol, *Genet. Sel. Evol.* **46**, 65 (2014).
47. S. Lepetz *et al.*, *J. Archaeol. Sci. Rep.* **40**, 103250 (2021).

48. G. K. Kunst, *Anthropozoologica* **31**, 109–118 (2000).
49. C. Johnstone, in *Feeding the Roman Army: The Archaeology of Production and Supply in NW Europe*, S. Stalibrass, R. Thomas, Eds. (Oxbow Books, 2008), pp. 128–145.
50. C. Gaunitz *et al*., *Science* **360**, 111–114 (2018).
51. S. Felkel *et al*., *Sci. Rep.* **9**, 6095 (2019).
52. F. B. Marshall, K. Dobney, T. Denham, J. M. Capriles, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6153–6158 (2014).
53. J. Lesur, E. A. Hildebrand, G. Abawa, X. Gutherz, *Quat. Int.* **343**, 148–158 (2014).
54. E. A. Hildebrand *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8942–8947 (2018).
55. K. MacDonald, R. Hutton MacDonald, in *The Origins and Development of African livestock: Archaeology, Genetics, Linguistics and Ethnography*, R. M. Blench, K. MacDonald, Eds. (University College London Press, 2000), pp. 8–9.
56. D. H. Alexander, K. Lange, *BMC Bioinformatics* **12**, 246 (2011).
57. N. Patterson, A. L. Price, D. Reich, *PLOS Genet.* **2**, e190 (2006).
58. N. Patterson *et al*., *Genetics* **192**, 1065–1093 (2012).
59. F. G. Vieira, A. Albrechtsen, R. Nielsen, *Bioinformatics* **32**, 2096–2102 (2016).
60. B. Q. Minh *et al*., *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

ACKNOWLEDGMENTS

Funding: L.K.H. thanks the director of the Tel es-Safi/Gath excavation, A. Maeir, for facilitating sampling. The DonkeyBank collection of modern donkey DNA samples is supported by the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 857251). L.P.P. is funded by national funds from FCT – Fundação para a Ciência e a Tecnologia, I.P. Genoscope and GeT-PlaGe sequencing platforms are partly funded by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09), and their "Grands Projets" funding instrument (BUCEPHALE). This project has received funding from the CNRS, University Paul Sabatier (AnimalFarm IRP), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreements 885729-AncestralWeave, 295729-CodeX, 853272-PALAEOFARM, and 681605-PEGASUS).
Author contributions: Conceptualization: L.O.; Materials and Reagents: J.-M.A., A.P., P.W., O.B., C.D., and L.O.; Archaeological samples and contextual information: M.V., E.V., M.Marzullo, O.P., U.T., G.B.G., A.T., V.T., F.A., J.L.C., M.J.V., M.T.A., N.B., A.G., L.K.H., A.Ö., B.S.A., V.O., B.C., S.L., A.A.V., H.D., A.M., M.Mashkour, and L.O.; Sampled modern donkeys: A.B.-P., D.-D.W., L.P.R., and S.A.B.; Radiocarbon dating: J.S.; DNA sequencing and preparation: L.T.C., L.C., S.S., A.F., A.S.-O., P.C., N.K., L.P.P., J.M.A., A.P., O.B., C.D., P.W., A.B.-P., and L.O.; Data analysis: E.T.T. and L.O. with input from P.L. and H.R.; Writing—Supplementary Information: E.T.T., with input from L.O.; Writing—Main Article: E.T.T. and L.O. with input from A.F., L.P.R., P.L., E.V., F.A., L.A.F., B.S., L.K.H., B.S.A., M.Mashkour, S.A.B., A.B.P., and all coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The sequence data generated in this study are available for download at the European Nucleotide Archive (accession no. PRJEB52849). The variant data for this study have been deposited in the European Variation Archive (EVA) at EMBL-EBI under accession number PRJEB55549. The recombination map can be downloaded from https://osf.io/k7x8b/. The accession numbers for each individual sample and all other data used in this study are included in tables S1, S2, and S11 of the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo3503
Materials and Methods
Figs. S1 to S15
Tables S1 to S11
References (*61*–*136*)
MDAR Reproducibility Checklist

Submitted 28 January 2022; accepted 15 August 2022
10.1126/science.abo3503

CANCER

Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling

Joseph M. Chan[1,2]†, Samir Zaidi[3,4]†, Jillian R. Love[3]‡§, Jimmy L. Zhao[3]‡¶, Manu Setty[2]‡#, Kristine M. Wadosky[5], Anuradha Gopalan[6], Zi-Ning Choo[2], Sitara Persad[2,7], Jungmin Choi[8], Justin LaClair[3], Kayla E. Lawrence[3], Ojasvi Chaudhary[2], Tianhao Xu[2], Ignas Masilionis[2], Irina Linkov[6], Shangqian Wang[3], Cindy Lee[3], Afsar Barlas[9], Michael J. Morris[4], Linas Mazutis[2,10], Ronan Chaligne[2], Yu Chen[3], David W. Goodrich[5], Wouter R. Karthaus[3]*§, Dana Pe'er[2,11]*, Charles L. Sawyers[3,11]*

Drug resistance in cancer is often linked to changes in tumor cell state or lineage, but the molecular mechanisms driving this plasticity remain unclear. Using murine organoid and genetically engineered mouse models, we investigated the causes of lineage plasticity in prostate cancer and its relationship to antiandrogen resistance. We found that plasticity initiates in an epithelial population defined by mixed luminal-basal phenotype and that it depends on increased Janus kinase (JAK) and fibroblast growth factor receptor (FGFR) activity. Organoid cultures from patients with castration-resistant disease harboring mixed-lineage cells reproduce the dependency observed in mice by up-regulating luminal gene expression upon JAK and FGFR inhibitor treatment. Single-cell analysis confirms the presence of mixed-lineage cells with increased JAK/STAT (signal transducer and activator of transcription) and FGFR signaling in a subset of patients with metastatic disease, with implications for stratifying patients for clinical trials.

Resistance to molecularly targeted therapies is often due to mutations in the drug target, but there is growing recognition of other modes of tumor escape, particularly involving next-generation inhibitors designed to circumvent target-based resistance mechanisms. For example, in epidermal growth factor receptor (EGFR)– and KRAS-mutant lung cancer and in metastatic prostate cancer, tumor cells can undergo a lineage transition from adenocarcinoma to squamous or neuroendocrine histology after treatment with the respective inhibitors (*1*–*6*). Similarly, BRAF-mutant melanomas can transition from a MITF$^+$ differentiated melanocyte phenotype to more mesenchymal or neural crest stem-like cell states in response to BRAF inhibition (*7*–*9*). Transcriptomic analyses indicate that these changing cell states share features of normal developmental, regenerative, and stem cell signatures, suggesting that tumor cells co-opt these gene expression programs to escape lineage-dependent therapies (*10*).

The increasing prevalence of this mode of tumor escape has sparked numerous investigations into the underlying mechanism. Analysis of patient specimens, particularly at the single-cell level, has been instructive in defining the complexity of the problem, revealing notable heterogeneity within and across patients (*11*, *12*). Because these studies typically capture a single snapshot during tumor evolution, it is difficult to discern precisely when, where, and how plasticity arises and its relationship with drug treatment and resistance.

Human prostate cancer provides an ideal model to study the role of plasticity in drug resistance because of its near universal association with prior exposure to androgen receptor (AR)–directed therapies. Prostate cancers initially present as AR$^+$ invasive adenocarcinomas that express luminal lineage markers, including CK8, CK18, and kallikreins such as KLK3 (prostate specific antigen or PSA). Histopathologic criteria for diagnosis classically include a loss of TP63$^+$ basal epithelial cells, underscoring the luminal-like identity of prostate cancer. Prostate cancers are uniformly sensitive to androgen deprivation therapy but eventually relapse as castration-resistant prostate cancer (CRPC).

Clinical observations and sequencing data demonstrate that CRPC has heterogeneous

[1]Department of Medicine, Thoracic Oncology Service, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [2]Program for Computational and Systems Biology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [3]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [4]Department of Genitourinary Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [5]Department of Pharmacology and Therapeutics, Roswell Park Cancer Institute, Buffalo, NY 14263, USA. [6]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [7]Department of Computer Science, Columbia University, New York, NY 10027, USA. [8]Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Korea. [9]Molecular Cytology Core Facility, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [10]Institute of Biotechnology, Life Sciences Centre, Vilnius University, Vilnius, Lithuania. [11]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.
***Corresponding author. Email: wouter.karthaus@epfl.ch (W.R.K.); peerster@gmail.com (D.P.); sawyersc@mskcc.org (C.L.S.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
§Present address: Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, EPFL, 1015 Lausanne, Switzerland.
¶Present address: AstraZeneca Oncology R&D, New York, NY 10016, USA. #Present address: Basic Sciences Division and Translational Data Science IRC, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

phenotypes (*13*). Two subtypes, adenocarcinoma (CRPC-adeno) and neuroendocrine prostate cancer (NEPC), can be broadly identified by histology and staining for luminal (AR, KLK2/3) and neuroendocrine (SYP, INSM1 CHGA/B, CD56) lineage markers. However, these markers do not capture the full spectrum of CRPC phenotypes, which include mixed luminal and neuroendocrine features, loss of canonical luminal and neuroendocrine markers ("double negative"), and gain of basal, mesenchymal, and stem-like gene expression signatures (*4*, *14*). Phenotypic heterogeneity in CRPC is thought to be a consequence of lineage plasticity within adenocarcinomas, rather than the outgrowth of preexisting subclones with alternative lineages. This conclusion is supported by the fact that genomic alterations found in early luminal cancers, such as ERG translocations, are retained in NEPC (*15*). Furthermore, these diverse CRPC phenotypes cannot be clearly distinguished by tumor genotype. Although the loss of TP53 and RB1 is highly enriched in NEPC and double-negative CRPC, alterations in these tumor suppressors, as well as PTEN, are also found in CRPC-adeno. A possible explanation is that loss of function in these genes may initiate a plastic cell state, which enables alternative lineage phenotypes to emerge during treatment with antiandrogen therapy.

Another critical question is whether anti-AR therapy, which is highly effective in inhibiting the growth of luminal epithelial tumor cells, may also accelerate the emergence of nonluminal CRPC phenotypes. The relative frequency of NEPC and double-negative CRPC cases has increased over the past decade (*4*), coinciding with increased clinical use of next-generation anti-AR therapeutics (such as abiraterone and enzalutamide) that inhibit AR signaling more potently than standard androgen-deprivation therapy. Although this correlation may be incidental, it begs the question of whether lineage plasticity, which is presumed to play a causal role in the development of antiandrogen resistance, is itself promoted or enhanced by anti-AR therapy.

Inflammatory signaling in adenocarcinoma cells precedes NEPC

To investigate these questions, we turned to two closely related genetically engineered mouse models (GEMMs) of prostate cancer that recapitulate the adenocarcinoma-to-NEPC transition after prostate-specific deletion of the human-relevant tumor suppressor genes *PTEN*, *RB1*, and *TP53* using CRE recombinase. Single-cell RNA sequencing (scRNA-seq) was performed on whole prostates harvested from 29 mice [9 wild type (WT); 7 $Pten^{-/-}Rb^{-/-}$ (PtR); 13 $Pten^{-/-}Rb^{-/-}Trp53^{-/-}$ (PtRP)] at relevant time points reflecting the adenocarcinoma-to-NEPC transformation (Fig. 1A and table S1) (*16*). Transcriptomes from 67,622 cells were analyzed and visualized by uniform manifold approximation and projection (UMAP), using coarse labeling to identify cell types and green fluorescent protein (*Gfp*) to mark cells that underwent recombination (Fig. 1B, fig. S1, and methods). In addition to the expected adenocarcinoma (*Epcam-Gfp*) and NEPC (Syp^+) populations, we were surprised to see three additional Gfp^+ clusters defined by expression of *Pou2f3* and *Dclk1* (*Pou2f3-Gfp*); *Vim*, *Twist2*, and *Ncam1* (*Vim-Gfp*); and *Tff3* (*Tff3-Gfp*), respectively. *Pou2f3* encodes a transcription factor (TF) expressed by rare chemosensory cells in normal gut and lung tissue (often called tuft cells) and defines a non-neuroendocrine variant subtype of small-cell lung cancer with morphologic features distinct from ASCL1 and NEUROD1 subtypes (*17*, *18*). *Tff3* encodes a secreted protein expressed by intestinal columnar epithelial cells involved in maintaining mucosal integrity and is a clinical biomarker for Barrett's esophagus (*19*).

We examined potential routes of plasticity by evaluating cell connectivity in k-nearest-neighbor graphs, visualized using force-directed layouts (FDLs) for each genotype across different time points. We noted the early appearance of Gfp^+ adenocarcinoma (*Epcam-Gfp*) as well as *Tff3-Gfp* cells, followed by the expansion of *Vim-Gfp* and NEPC and *Pou2f3-Gfp* populations, with the transition to NEPC more prevalent in the PtRP genotype, indicative of the additional impact of P53 loss (Fig. 1, C and D). The presence of each of these subpopulations was confirmed by multiplex immunofluorescence (IF) and immunohistochemistry (IHC) (table S2), which documented SYP^+ (NEPC) and VIM^+ (VIM) cells in a background of $ECAD^+$ (and $PANCK^+$) epithelial cells (top row for ECAD in Fig. 1E and PANCK stain in fig. S2) as well as larger clusters of NEPC (top row) and VIM (bottom row) and foci of $TFF3^+$ or $POU2F3^+$ cells (bottom rows) (Fig. 1E). Although this unexpectedly diverse range of cell types, often seen within a single mouse, is of interest, we focused our analysis here on the adenocarcinoma and NEPC populations because of their established clinical importance. Deeper analysis of the $Tff3^+$ and $Pou2f3^+$ populations will be reported separately.

Owing to the large number of independent alleles, definitive lineage-tracing experiments cannot be easily performed in this GEMM. Consequently, we relied on various computational approaches to predict the most likely scenario by which lineage plasticity emerges. As a starting point, we reasoned that the early emergence of the adenocarcinoma population, along with its placement between WT and other mutant clusters in the k-nearest-neighbor graph, implicate it as the most likely source of plasticity (fig. S3, A to D, and methods). Therefore, we examined differentially expressed genes (DEGs) and pathways associated with populations along this putative transition, using gene set enrichment analysis (GSEA) (Fig. 1F and tables S3 to S5). First, we noted enrichment for a stem-like (L2) luminal signature that we and others have previously reported from single-cell analysis of the normal mouse prostate (*20*–*22*) as well as enrichment for epithelial-mesenchymal transition (EMT) programs and for signatures of early developmental programs (*23*, *24*). Most surprising was strong enrichment for a range of inflammatory response gene signatures, including *Ifn-a*, *Ifn-g*, *Jak/Stat*, and *Tnfa* (table S4). Given the indications implicating the adenocarcinoma population as a potential source of plasticity and its clinical relevance, we used Palantir trajectory analysis (*25*) to consider potential routes between adenocarcinoma and NEPC (Fig. 1G and methods). We noted three major trends in TF expression: (i) those active in adenocarcinoma alone, (ii) those expressed during the putative transition to NEPC, and (iii) NEPC-specific regulators (see Fig. 1, G and H, and fig. S3, E and F, for robustness analysis and table S5). Notably, four inflammatory response TFs (Stat1, Stat2, Irf7, and Irf1), along with JAK/STAT (Janus kinase/signal transducer and activator of transcription) signaling, are specifically enriched in adenocarcinoma cells during the transition phase to NEPC (Fig. 1, G and H).

Cell-autonomous induction of plasticity in normal prostate organoids

Having documented the onset and evolution of lineage plasticity in vivo in the mouse prostate, we next asked whether plasticity is dependent on the tumor microenvironment or whether it can occur in a cell-autonomous manner. For this purpose, we turned to organoid culture because prostate epithelial cells can be grown in isolation without accessory stromal or immune cell populations from the prostate microenvironment (*26*). Because this system supports the growth of normal as well as malignant epithelial cells, it also provides an opportunity to examine the immediate and longer-term consequences of tumor suppressor gene deletion with precise kinetics.

Focusing initially on *Rb1* and *Trp53* deletion (to model the alterations in human NEPCs in a minimalist system), we infected $Trp53^{loxP/loxP}$; $Rb1^{loxP/loxP}$ organoids with Cre-expressing lentivirus. We also leveraged a mTmG reporter cassette to identify $Trp53^{-/-}$;$Rb^{-/-}$ cells using the red-to-green color change initiated by Cre expression (movie S1). In contrast to the cystic appearance of uninfected organoids, those with *Trp53* and *Rb1* codeletion lost their basal-luminal polarity and cystic lumens by ~4 weeks (hereafter called hyperplastic), followed by the appearance of migratory cells and cellular protrusions outside the normally sharp exterior organoid borders at ~8 to 10 weeks (hereafter called "slithering," based on similarity to

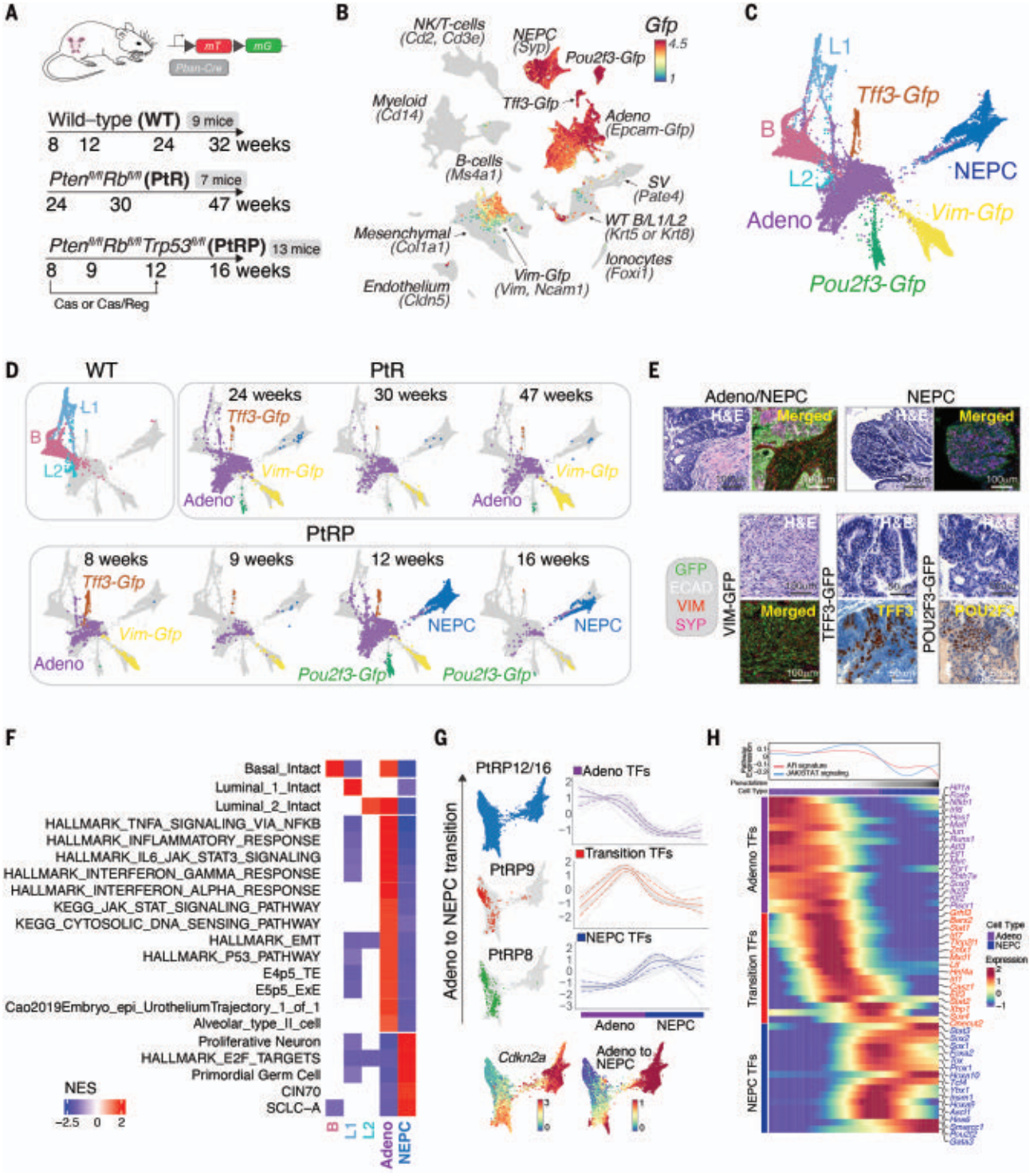


Fig. 1. Inflammatory JAK/STAT signals emerge in adenocarcinoma before transition to NEPC in GEMMs. (**A**) Experimental design includes WT (9 mice), PtR (7 mice, *PbCre:Rosa26*$^{mT/mG}$*Pten*$^{fl/fl}$*Rb1*$^{fl/fl}$), and PtRP (13 mice, *PbCre: Rosa26*$^{mT/mG}$*Pten*$^{fl/fl}$*Rb1*$^{fl/fl}$*Trp53*$^{fl/fl}$) at labeled time points. Time points for WT included 8 (one mouse), 12 (two mice), 24 (three mice) and 32 (three mice) weeks; for PtR included 24 (three mice), 30 (two mice), and 47 (two mice) weeks; and for PtRP included 8 (one mouse), 9 (one mouse), 12 (four mice), and 16 (one mouse) weeks relevant to the adenocarcinoma-to-neuroendocrine transition (*4*–*6*). Furthermore, six PtRP mice were castrated at 8 weeks of age for either a total of 4 weeks (three mice, "Cas") or 2 weeks followed by 2 weeks of dihydrotestosterone (DHT, "Cas/Reg") addback (three mice). (**B**) UMAP of GEMMs (N = 67,622 cells) shown for all cell types based on imputed *Gfp* expression (restricted to nonimmune cells) as a marker for mutant cells (see methods). (**C**) FDL of mutant *Gfp*$^{+}$ and WT cells (N = 28,934) colored by cell type. WT includes luminal 1, luminal 2, and basal. Mutant cell types include adenocarcinoma, *Tff3-Gfp*, *Pou2f3-Gfp*, *Vim-Gfp*, and NEPC. (**D**) FDLs separated by genotype and time point and colored by cell type: WT with N = 7435 cells (top left); PtRP 24, 30, and 47 weeks with N = 2565, 662, and 1441 cells, respectively (top right); and PtRP 8, 9, 12, and 16 weeks with N = 981, 353, 4984, and 569 cells, respectively (bottom). (**E**) Hematoxylin

and eosin (H&E) and multiplex immunofluorescence (mIF) (40×) of mutant subtypes, including adenocarcinoma-to-NEPC transition, NEPC, and VIM. mIF channels include GFP (mG, green), 4′,6-diamidino-2-phenylindole (DAPI) (nuclei, blue), VIM (mesenchymal, red), SYP (synaptophysin, purple), and ECAD (e-cadherin, white). Also shown are H&E and IHC of TFF3$^+$ and POU2F3$^+$ cells in mutant tissue. Scale bars represent 100 or 50 μm as noted. (**F**) Heatmap of significantly enriched gene sets per cell type, restricted to WT epithelial (basal, luminal 1, luminal 2), adenocarcinoma, and neuroendocrine (NEPC) cells [FDR < 0.01, abs(NES) > 1 where NES is the normalized enrichment score; see methods]. (**G**) Dissecting the adenocarcinoma-to-NEPC transition. Shown on the left are a FDL of mutant adenocarcinoma (subsetted to adeno-B and adeno-L2) and NEPCs cells from the PtR model (N = 16,593 cells) colored by genotype and time. Shown at the bottom are FDLs colored by normalized log2(X + 1) expression of *Cdkn2a* or pseudotime scaled from 0 to 1. Shown on the right are gene trends for TF DEGs across the adenocarcinoma to NEPC branch probability that were calculated using a generalized additive model, with cubic splines across 500 equally sized bins (solid lines represent the mean; dashed lines represent the standard deviation) (*25*). Gene trends were grouped by PhenoGraph cluster (k = 30) (see methods). (**H**) TF groups include adenocarcinoma (purple), putative transition (red), and NEPC (blue). Heatmap of gene trends of select TF DEGs (PhenoGraph cluster versus the rest within PtRP; refer to methods section "Differential expression in single-cell data") from each category are ordered by the putative transition from adenocarcinoma to NEPC, with gene labels colored in the same way the aforementioned groups (scale gene trends of imputed expression, −0.5 to 1.5). The top panel shows a spline fit of the average Z-score of JAK/STAT and androgen signaling (NELSON_ANDROGEN_SIGNALING_UP) (see methods). Adeno, adenocarcinoma; B, basal; L1, luminal 1; L2, luminal 2; SV, seminal vesicles; Vim-Gfp, Vimentin-Gfp.

the migratory phenotype originally reported for lung neuroepithelial bodies) (*27*) (Fig. 2, A and B; figs. S4, A to C, and S5; and movie S2). This phenotype was also seen after CRISPR-Cas9–mediated codeletion of *Trp53* and *Rb1* in WT organoids (fig. S4B) and required deletion of both tumor suppressors. No morphologic changes were observed after deletion of *Trp53* or *Rb1* alone (fig. S4B), consistent with earlier work in human prostate cancer cell lines (*28*).

To further characterize the changes induced by *Trp53* and *Rb1* codeletion, we performed bulk RNA-seq, Western blot analysis, and IF or IHC analysis of various lineage markers. The most dramatic findings were reduced expression of luminal lineage genes such as *Nkx3.1*, *Dpp4* (*Cd26*), *Krt18*, and *Krt8* and increased expression of mesenchymal genes such as *Snai1*, *Zeb2*, and *Vim* (Fig. 2, C to E; fig. S4, C to E; and table S6). GSEA analysis of the bulk RNA-seq data confirmed enrichment of EMT (Fig. 2F and table S7) as seen in PtRP mice, which was further supported by increased protein levels of Slit3 (also in the leading edge of the EMT gene set) and the migratory slithering phenotype observed morphologically.

Most surprising, however, was strong enrichment for the same inflammatory response pathways (e.g., Jak/Stat) that we saw earlier in PtRP GEMMs, as well as overexpression of *Fgfr1* [log2 fold change of 5.0; false discovery rate (FDR) < 7×10^{-199}] and the ligands *Fgf1*, *Fgf2*, *Fgf7*, and *Fgf10* (all fold change >4; FDR < 5×10^{-8}) (Fig. 2, E and F). We confirmed this activation biochemically, with robust pStat1 and pStat3 activation in *Trp53*$^{-/-}$;*Rb1*$^{-/-}$ organoids that was reversed by 48-hour treatment with the Jak kinase inhibitor ruxolitinib (Rux) (Fig. 2, F and G, and fig. S4F). Thus, the early manifestations of plasticity seen in GEMMs can be recapitulated in normal prostate epithelial cells in culture in the absence of any accessory cells, indicative of a cell-autonomous process.

However, unlike the later NEPC phenotype observed in PtRP mice and previous reports of *Trp53*$^{-/-}$;*Rb1*$^{-/-}$ GEMMs (*29*), we failed to detect pronounced expression of *Syp* or other neuroendocrine markers (*Chga*, *Chgb*, *Ascl1*, or *Neurod1*), even in *Trp53*$^{-/-}$;*Rb1*$^{-/-}$ organoids passaged for >12 months. Postulating that this might be due to the absence of critical prostate microenvironmental factors in organoid culture media, we asked if the organoid phenotype evolved further after orthotopic (OT) implantation of *Trp53*$^{-/-}$;*Rb1*$^{-/-}$ organoids into *Trp53*$^{+/+}$;*Rb1*$^{+/+}$ hosts. After confirming highly efficient engraftment of hyperplastic and/or slithering stage organoids in a pilot experiment (by detection of GFP$^+$ cells at 4 weeks in the dorsal lobe of hormonally intact NSG mice), we scored lineage phenotype and tumorigenicity, as well as the effect of androgen blockade, in a larger cohort by treating half the mice with castration and daily enzalutamide (Enz) 4 weeks after OT injection. In the hormonally intact mice, engrafted cells remained Ar$^+$, Syp$^-$ (as in organoid culture); however, large foci of Ar$^-$, Syp$^+$ engrafted cells were seen in all mice in the castration with Enz group, indicating progression to a neuroendocrine-like lineage, which was a consistent finding across four independent *Trp53* and *Rb1* codeletion experiments (fig. S6, A to D). To determine if this effect of in vivo androgen blockade on lineage plasticity is also observed in culture, we returned to the organoid model to determine the effect of Enz on the slithering phenotype after *Trp53* and *Rb1* codeletion. The percentage of organoids scored as slithering increased from ~10 to >30% (~3-fold) after 7 days of Enz treatment (fig. S6E).

Collectively, the above experiments demonstrate that changes in lineage-marker expression and inflammatory response pathway activation seen in the adenocarcinoma population in PtRP mice can be recapitulated in mouse prostate organoids within weeks of *Trp53* and *Rb1* codeletion. Furthermore, anti-AR therapy (castration and/or enzalutamide) can accelerate the rate at which these changes evolve, consistent with recent data using a patient-derived xenograft model (*30*). To address which cells can give rise to the plasticity phenotypes described above, we took advantage of prior work showing that organoids with normal basal and luminal bilayer architecture can be generated from single luminal or basal cells (*26*). Basal (CD49f$^+$) and luminal (CD24$^+$) cells from established Rosa26-lsl-Cas9 organoids stably expressing guide RNAs targeting *Trp53* and *Rb1* were isolated by fluorescence-activated cells sorting (FACS), transduced with adenoviral Cre, seeded as single cells, expanded, and then scored after 4 weeks using the morphologic criteria for plasticity (cystic versus hyperplastic, described in Fig. 2B). About 80% of plated basal and luminal cells developed hyperplastic changes, indicating that either cell type has lineage plasticity potential in organoid culture (Fig. 2H and fig. S4, H and I).

Lineage plasticity initiates with mixing of basal-luminal identity

Our organoids constitute a tractable experimental model that initiates and elaborates lineage plasticity in a reproducible manner. To gain more molecular insight into this plasticity, we characterized emerging transcriptional cell states and gene programs by collecting scRNA-seq data, beginning 2 weeks after *Trp53* and *Rb1* codeletion (before any morphologic evidence of plasticity) as well as 4 and 8 weeks after deletion. A parallel set of cultures were treated with Enz at 2 weeks and harvested at 4 and 8 weeks (Fig. 3A and table S1) to characterize the effects of androgen blockade on plasticity (see fig. S6).

We observed great transcriptional heterogeneity both within and between each time point and condition. To systematically catalog the observed cell states, we started with the WT phenotypic landscape then followed how it remodels after perturbation. To gain a concise description of the observed phenotypic diversity, we used our SEACells algorithm (*31*) (see methods) to calculate metacells separately within each sample. The metacell approach (*24*) aggregates similar single cells into a set of distinct, highly granular cell states such that any differences between cells within a metacell are likely technical rather than biological. Aggregating counts within each metacell generates a robust (fig. S7), full transcriptomic

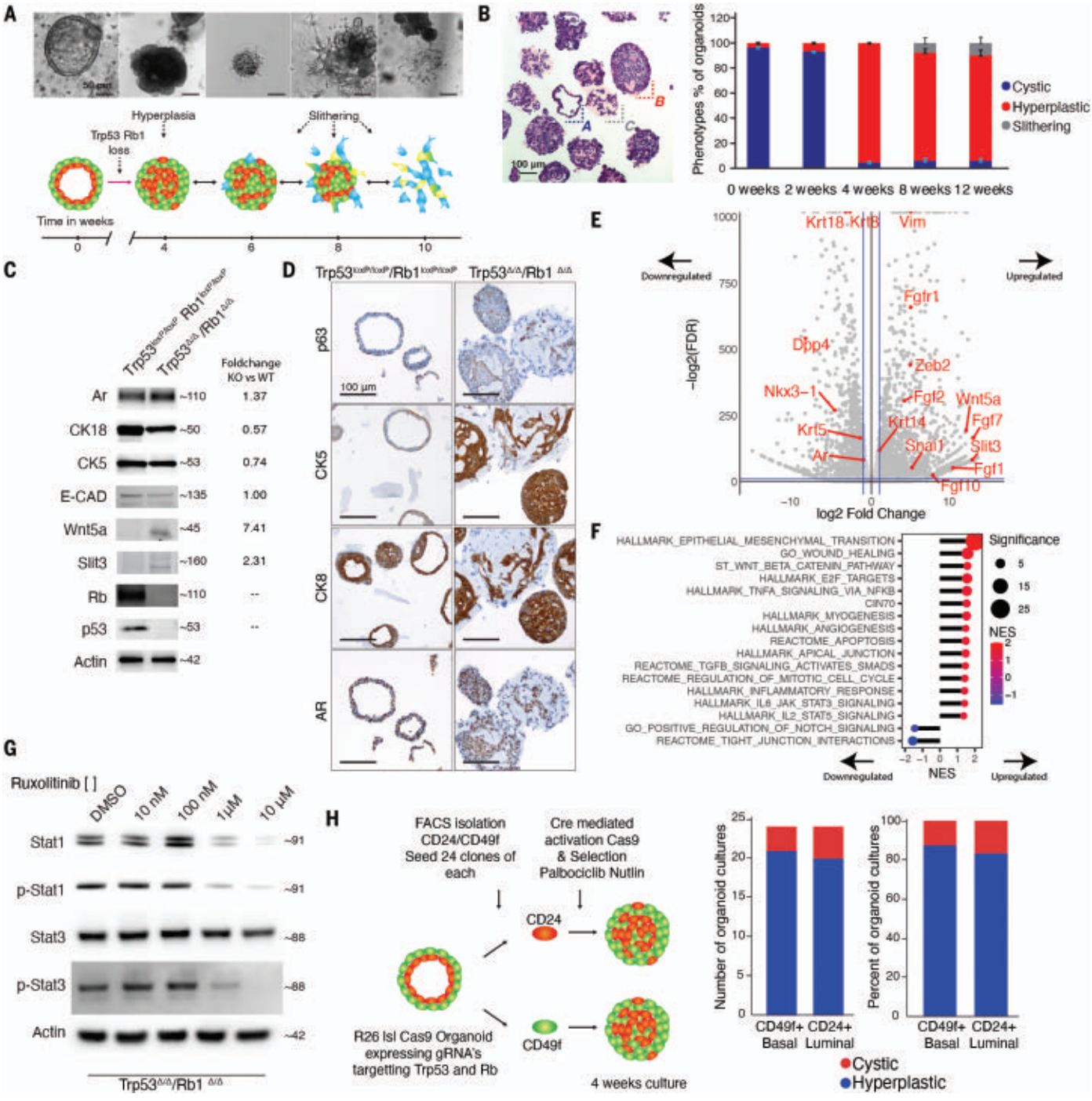


Fig. 2. Establishment of an organoid model of spontaneous lineage plasticity. (**A**) At the top are representative brightfield microscopy images of the various organoid phenotypes before and 4, 6, 8, and 10 weeks after *Trp53* and *Rb1* loss. At the bottom is a schematic representation of lineage plasticity development in $Trp53^{\Delta/\Delta};Rb1^{\Delta/\Delta}$ organoids. Scale bars represent 50 µm. (**B**) On the left is a representative H&E staining of LentiCre $Trp53^{\Delta/\Delta};Rb1^{\Delta/\Delta}$ organoids ~10 weeks after deletion with cystic (A, blue), hyperplastic (B, red), and slithering (C, silver) phenotypes. On the right is a bar graph showing the percentage of organoids with given phenotypes during the time course. The scale bar represents 100 µm. (**C**) Western blot verification of DEGs identified in RNA-seq data in WT ($Trp53^{loxP/loxP};Rb1^{loxP/loxP}$ organoids, left) and ~10 weeks after deletion (LentiCre $Trp53^{\Delta/\Delta};Rb1^{\Delta/\Delta}$ organoids, right). Proteins are as marked. Actin was used as a loading control. Fold change in protein level as determined with ImageJ is given on the right. KO, knockout. (**D**) Representative IHC of basal markers (p63, Ck5), luminal marker (Ck8), and AR in WT ($Trp53^{loxP/loxP};Rb1^{loxP/loxP}$ organoids, left) and ~10 weeks after deletion (LentiCre $Trp53^{\Delta/\Delta};Rb1^{\Delta/\Delta}$ organoids, right). Scale bars represent 100 µm. (**E**) Volcano plot showing DEGs in the mutant organoid compared with WT. Blue lines indicate thresholds for significant DEGs [Bonferroni-adjusted *p* value < 0.001, abs(log2 fold change) > 1]. Red labels indicate significant DEGs of interest. (**F**) Lollipop plot showing significantly enriched pathways in *Trp53* and *Rb1*–deleted organoids versus WT using bulk RNA-seq based on GSEA [Benjamini-Hochberg adjusted *p* value < 0.15; abs(NES) > 1] (red indicates enriched in mutant; blue indicates enriched in WT). Lollipop size corresponds to significance, or −log2(*p* value). (**G**) Western blot of total Stat1 and Stat3 levels and phosphorylated (p-) Stat1 and p-Stat3 in LentiCre $Trp53^{\Delta/\Delta};Rb1^{\Delta/\Delta}$ organoids treated with increasing doses of the JAK1/JAK2 inhibitor Rux for 48 hours. Actin was used as a loading

control. (**H**) On the left is a schematic representation of single-cell cloning experiment of basal (CD49f$^+$) and luminal (CD24$^+$) cells. Organoids harboring a Cre recombinase–inducible Cas9 [Rosa26 lsl-Cas9 (*32*)] were transduced with guide RNAs targeting *Trp53* and *Rb1* (LentiGuide Puro, guide sequences; table S23). Basal and luminal cells were sorted based on CD49f$^+$ (basal) and CD24$^+$ (luminal) expression, and subsequently, Cas9 was activated by transduction with adenovirus expressing Cre. About 1000 single cells were seeded and grown out. Nutlin-3 and Palbociclib were added 3 days after Cas9 activation to select for *Trp53* and *Rb1* mutant cells. Twenty-four basal and 24 luminal clones were randomly chosen and expanded for 4 weeks. On the right is a quantification of whole-culture phenotypes (cystic or hyperplastic) of single cell–derived (CD49f$^+$ or CD24$^+$ FACS sorted) organoids after 4 weeks of culture. The left bar graph shows the absolute number of cultures, and the right bar graph shows the percentage of single-cell derived cultures in a cystic or hyperplastic state.

quantification for each cell state, thus mitigating noise and sparsity in scRNA-seq data. An added benefit of metacells is the enumeration of observed cell states that are more amenable to comparison.

To establish a baseline for assessing plasticity changes in mutant organoids, we first focused on the cell states in WT organoids. PhenoGraph clustering (*25*) identified five distinct subpopulations, which we classified as luminal (L-Org1, L-Org2) or basal (B-Org1, B-Org2, B-Org3) based on published transcriptomes (see methods) (Fig. 3, B and C; fig. S8, A and B; and tables S8 and S9). L-Org1 cells share expression of stem-like genes (e.g., *Ly6d*, *Tacstd2/Trop2*) with L-Org2 cells but also express other cytokeratins (e.g., *Krt6a*, *Krt10*) not typically associated with luminal cells. L-Org2 cells resemble stem-like (*Ly6a/Sca1*$^+$, *Psca*$^+$) luminal cells seen in vivo [called L2 in Karthaus *et al.* (*20*); LumP in Crowley *et al.* (*21*)] as well as canonical luminal cytokeratins (*Krt8/Krt18*) (*20*, *21*). B-Org1 cells express canonical basal markers (*Krt5*, *Krt14*, *Trp63*) and B-Org2 cells express secretory proteins such calcitonin and inhibin A, whereas B-Org3 cells are largely distinguished by expression of proliferation markers (*Mki67*, *Aurka*).

After analyzing the entire time course after *Trp53* and *Rb1* codeletion, we found no overlap between cell states in WT and mutant organoids, even at 2 weeks, when organoids were still morphologically normal. When visualized in a FDL (see methods), these metacells organize along the axis of time with a distinct gap between the WT and 2-week time point. Basal and luminal metacells identified by correlation to basal and luminal gene signatures are phenotypically well separated in the WT setting but converge together on the FDL after mutation (Fig. 3D and fig. S8, C and D; see fig. S8E for similar trends at the single-cell level). To understand these phenotypic changes, we characterized the basal and luminal features gained and lost after mutation (tables S10 to S12; see methods). Over time, mutant organoids gain a mixed phenotype, in which mutant basal metacells acquire typical luminal features (*KRT8*, *KRT18*, *PROM1*) and lose typical basal features (*KRT5, KRT14*). In parallel, mutant luminal metacells gain typical basal features (*KRT17*) but lose luminal markers (*PSCA*, *LY6D*) (fig. S8, F and G). At the same time, both basal and luminal mutant cells concurrently gain EMT-like features (*VIM*, *FN1*, *COL4A1/2*, *ZEB1/2*). Plasticity changes in the mutant organoid can thus be defined as a gain in mixed basal-luminal phenotype with EMT-like features. Furthermore, all of these changes are escalated by antiandrogen (Enz) treatment, consistent with the accelerated morphologic phenotypes observed in culture and in vivo.

To ensure that the observed mixed basal-luminal phenotype was not merely a loss of original basal-luminal identity or a product of the metacell approach, we calculated basal and luminal scores for each individual cell based on basal and luminal features gained after mutation (table S12; see methods). Corresponding density plots showed that mutant luminal cells become more basal-like over time, and vice versa, resulting in a convergence between luminal and basal phenotype over time and with treatment (Fig. 3E and fig. S9A).

The apparent infidelity in basal-luminal lineage suggested that the mixed phenotype could serve as a useful proxy for quantifying plasticity. We therefore developed alternative measures of basal-luminal mixing at each time point as a proxy for plasticity (see methods). These per-sample metrics confirmed that plasticity increases over time after tumor suppressor deletion and with Enz treatment (Fig. 3F and fig. S9). As a per-cell measure that captures plasticity changes both within and across sample time points, we used a cell-type classification-based approach (*25*) to associate each mutant cell with a vector of classification probabilities for each WT organoid phenotype, where classification uncertainty corresponds to increased mixing and plasticity (see methods). We visualized these per-cell probabilities as a ternary plot (fig. S10, A and B). As expected, WT cells favor a single assignment to their respective phenotype, whereas mutant cells converge to maximal uncertainty at the center of the ternary plot over time. We quantified this time-dependent increase in basal-luminal mixing using entropy of cell-type probabilities, which robustly agrees with our per-sample measures of plasticity while capturing a range of plasticity within each time point and condition (fig. S10, C and D; see methods). Interestingly, Enz-treated samples display even fewer cells favoring the L-Org2 phenotype, consistent with our transcriptional and morphological data showing that AR inhibition accelerates the loss of luminal identity (figs. S6D and S10E). Notably, we observed similar basal-luminal mixing in the PtRP mice, further validating the use of the organoid platform for interrogating plasticity phenotypes (fig. S10, F to J). Thus, several orthogonal analytical approaches consistently detect increasing plasticity over time, in both organoids and GEMMs.

Jak/Stat and Fgfr up-regulation coincides with plasticity onset and progression

To uncover the gene programs that drive plasticity after tumor suppressor gene deletion, we leveraged our per-cell measure of plasticity to identify pathways activated in highly plastic cells. After correlating genes with our plasticity measure, we used GSEA to identify gene programs significantly associated with plasticity (fig. S11A and tables S13 and S14; see methods). Among these plasticity-associated pathways, we noted up-regulated EMT and NEPC signatures (fig. S11, B and C), even though morphological evidence of neuroendocrine transformation was not observed until these organoids were studied in vivo (fig. S6, A to D). To enrich for causal genes, we narrowed our list of candidate driver programs to pathways that meet two additional criteria: (i) activated early in the time course, because plasticity emerges before morphological changes (week 2 after *Trp53* and *Rb1* codeletion) and (ii) further enriched with Enz treatment, based on the premise that plasticity increases with therapeutic pressure (tables S15 and S16; see methods). The top resulting pathways include JAK-STAT [also leukemia inhibitory factor (LIF), interferon-γ (IFN-γ), IFN-α], fibroblast growth factor receptor (FGFR), NOTCH, and EMT (also SMAD2) (Fig. 3G and fig. S11, D to F).

Because our prostate organoids are cultured using defined, serum-free conditions that contain no ligands that might activate IFN response genes, we searched for potential mechanisms to explain the IFN response. Increased IFN signaling is often a consequence of cytoplasmic sensing of double-stranded DNA or RNA that is due to ruptured micronuclei or reactivation of endogenous retroviruses (*32*). However, CRISPR deletion of several genes involved in this sensing pathway [*Tmem173* (Sting), *Mavs*, *Tbk1*, *Ifih1*] had no effect on the acquisition of the hyperplastic morphology after *Trp53* and *Rb1* codeletion. We therefore explored the possibility of an autocrine loop, reasoning that if a ligand-receptor (L-R) pair activates JAK/STAT, then high abundance of the L-R should correlate with increased JAK/STAT signaling in the mutant organoid cells. A key advantage of our

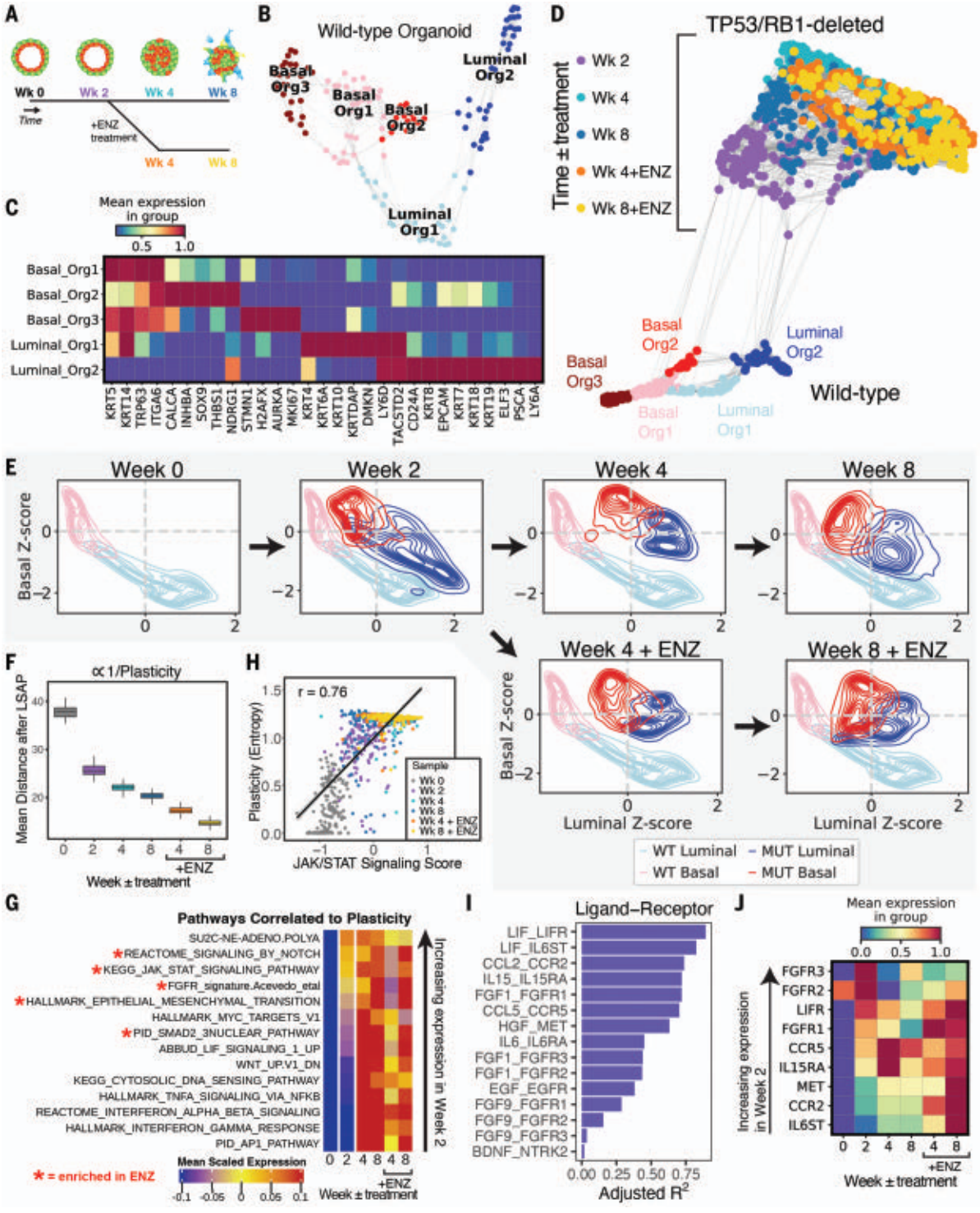


Fig. 3. Single-cell analysis of organoids reveals basal-luminal mixing, used as a proxy measure for plasticity to identify associated inflammation, JAK/STAT, and FGFR pathway activation. (**A**) Experimental design of organoid sequencing time course. Samples collected at weeks 0, 2, 4, and 8 are enzymatically digested and subjected to scRNA-seq. Samples from weeks 4 and 8 are collected with or without enzalutamide (ENZ) treatment. (**B**) FDL of WT basal and luminal SEACell metacells (N = 142), labeled by WT subpopulations determined by PhenoGraph clusters. Edges between metacells indicate the k-nearest neighbors (k = 6) (see methods). (**C**) Mean log2(X + 1) expression of DEGs in each basal and luminal subpopulation in the WT organoid. (**D**) FDL (k = 10) of metacells in the prostate organoid before and after *Trp53* and *Rb1* deletion (N = 884), annotated by either WT subpopulations or by time and treatment after mutation. (**E**) Contour plots of basal and luminal cell densities at each time point, depicting a convergence of cell identities after mutation (see methods). Single cells are plotted using mean Z-scores of the set of basal genes that are gained in mutant luminal cells (y axis) and the set of luminal genes that are gained in mutant basal cells (x axis). Samples were collected at week 0 (N = 2629), 2 (N = 1904), 4 (N = 2554), 8 (N = 2690), 4+ENZ (N = 2850), and 8+ENZ (N = 3050). Gene imputation is performed using MAGIC (k = 30; t = 3) before obtaining the Z-score. See fig. S9A for a corresponding version without

gene imputation. Red, basal cell; blue, luminal cells; light colors, WT; dark colors, *Trp53* and *Rb1* mutant (MUT). (**F**) Mean Euclidean distance between matched basal and luminal cells, based on the linear sum assignment problem (LSAP). Mean distances within each time point are an inverse measure of plasticity, which increases with time and treatment. (**G**) Top pathways significantly enriched for plasticity across time points (using GSEA, Bonferroni-adjusted $p < 0.01$, NES > 1; see fig. S10C and table S14). Each pathway score is measured as the average *Z*-score of gene expression in each pathway among metacells. Rows (pathways) are ordered by increasing pathway score from bottom to top in the early time point at week 2. Red asterisks indicate pathways that are also enriched in ENZ-treated samples compared with untreated samples (see methods). (**H**) Cellular plasticity increases as a function of JAK/STAT signaling. Plasticity is measured as the average entropy of cell-type classification probabilities per metacell. JAK/STAT score is the average *Z*-score of the leading-edge gene subset of KEGG_JAK_STAT_SIGNALING_PATHWAY and HALLMARK_IL6_JAK_STAT3_SIGNALING. The entropy of classification probabilities was first calculated for each single cell and then averaged per metacell for visualization. A linear fit is shown with a Pearson's correlation of 0.76. (**I**) Top-scoring L-R interactions known to activate JAK/STAT signaling, based on the adjusted coefficient of determination (R_{adj}^2) of the regression *JAK_STAT* ~ *L* + *R* + *L:R*, where *JAK_STAT* is the JAK/STAT signaling score and *L* and *R* represent scaled imputed ligand and receptor expression (see methods). Only L-R pairs with nonzero R_{adj}^2 are shown. (**J**) Mean expression (rows) by sample time point (columns) for candidate receptors that activate downstream JAK/STAT signaling in metacells. Receptor genes are ordered by increasing expression in the early time point at week 2 from bottom to top.

organoid model over the in vivo PtRP GEMM model is that it only needs to consider autocrine L-R interactions; thus, we were able to exclude cell type as a variable in L-R analysis. We evaluated a set of 74 L-R pairs known to activate JAK/STAT signaling and scored each candidate based on their correlation with a summary statistic of 33 JAK/STAT-associated genes enriched after mutation (tables S17 and S18; see methods).

The top L-R pairs identified by this analysis were LIF and its heterodimeric receptor (LIFR/IL6ST), FGF1/FGFR, IL15/IL15RA, the chemokines CCL2 and CCL5 and their cognate receptors CCR2 and CCR5, and HGF/MET (Fig. 3I and fig. S12A). To prioritize L-R pairs that cause an increase in plasticity, we applied similar criteria as for our candidate driver programs: (i) up-regulation of receptor expression early in the time course (week 2) and (ii) further up-regulation with antiandrogen treatment. FGFR and LIF (co)receptors remained as top candidate drivers after filtering, based on increased expression of FGFR1, LIFR, and IL6ST after ENZ treatment as well as early expression of FGFR1-3 and LIFR after *Trp53* and *Rb1* codeletion (Fig. 3J and fig. S12, B and C). Thus, multiple lines of evidence from PtRP GEMM mice and from *Trp53* and *Rb1*–deleted organoids implicate Jak/Stat pathway activation as an early event that precedes the acquisition of various plasticity phenotypes such as basal-luminal lineage mixing, EMT, and transition to NEPC, with FGFR and LIF prioritized as top candidates for functional interrogation based on L-R analysis.

Jak/Stat and Fgfr inhibition restores luminal lineage identity

To test these predictions at a functional level, we added recombinant FGF1 or LIF to organoids after *Trp53* and *Rb1* codeletion, reasoning that plasticity might be further enhanced by an artificial increase in ligand concentration. We also included hepatocyte growth factor (HGF), which was among the top candidates in our L-R analysis but deprioritized because of its delayed activation relative to the onset of the plasticity phenotype and thereby acted as a validation control for our prioritization schema. Recombinant FGF1, but not LIF or HGF, enhanced the magnitude of slithering (from 6 to 50% of Matrigel surface) within 3 days (fig. S13A; arrows indicate slithering), providing functional support for Fgfr signaling as a driver of plasticity.

To address the functional role of JAK/STAT signaling, we derived $Trp53^{loxP/loxP}$, $Rb1^{loxP/loxP}$ organoids in which either *Jak1* or *Jak2* alone or in combination were disrupted using CRISPR-Cas9–directed single guide RNAs. We then deleted *Trp53* and *Rb1* by lentiviral Cre infection and scored the percentage of organoids with cystic, hyperplastic, or slithering phenotypes at 8 weeks, as in Fig. 2B. Disruption of *Jak1* or *Jak2* alone blunted up-regulation of pStat1 and pStat3 but only modestly affected development of the plasticity phenotype [~6 to ~15% increase in cystic organoids ($p = 0.01$, Jak1 guides; $p = 0.09$, Jak2 guides)], with most organoids progressing to the hyperplastic morphology (fig. S13, B and C). By contrast, organoids with combined Jak1/Jak2 disruption (which more effectively blocked Stat1/3 phosphorylation) had substantial preservation of cystic morphology (>3-fold increase; fig. S13, B and C). We confirmed this effect pharmacologically by treatment with the dual Jak1/2 kinase inhibitor Rux, which restored cystic morphology (from ~5 to ~30%) in a dose-dependent manner (Fig. 4A and fig. S13, D and E). As further evidence of reversion to a more luminal lineage, these morphological changes were accompanied by a modest reduction in Vim^+ cells and an increase in AR expression (Fig. 4, C and D), commensurate with the reduction of pStat1 and pStat3 levels back to baseline.

Having demonstrated that perturbation of either FGFR or JAK/STAT signaling can affect the plasticity phenotypes in our organoid model, we used the clinical grade FGFR inhibitor erdafitinib (Erda) to examine the functional consequences of pharmacological FGFR inhibition on plasticity, alone and in combination with JAK inhibition. Although treatment with Erda also reduced the percentage of organoids with the slithering phenotype, we observed a notable ~12-fold increase (5 to 60%) in cystic morphology when fully plastic organoids were cotreated with Erda and Rux (Fig. 4, A and B).

Consistent with the morphologic change, we noted >3-fold reduction in cells expressing the mesenchymal lineage marker Vim by IHC (Fig. 4C), which was confirmed by Western blot and increased levels (1.5-fold) of the luminal markers Ck8 and AR (Fig. 4D and fig. S13F). Importantly, these changes in morphology and lineage-marker expression occurred in the absence of any effects on organoid proliferation (Fig. 4E and fig. S13G), suggesting that the primary consequence of combined Jak/Fgfr inhibition is on lineage reprogramming rather than selective outgrowth of rare organoids that had retained cystic morphology after *Trp53* and *Rb1* codeletion. The restoration of cystic luminal organoids after combined Fgfr with Jak kinase inhibition led us to ask if sensitivity to antiandrogen therapy is also restored. Indeed, Enz sensitivity was restored to previously Enz-resistant organoids (~50% decrease in proliferation, comparable to WT organoids), but only after cystic morphology was restored by 14 days of Rux with Erda pretreatment (Fig. 4E and fig. S13H).

JAK/STAT and FGFR activation in human CRPC

To determine if the transcriptional programs linked to plasticity in our murine prostate models are used in human prostate cancer, we analyzed scRNA-seq data from 12 CRPC patient samples representing both CRPC-adeno and transformed NEPC histologies (fig. S14 and table S19) that were obtained by radiologically guided biopsies of metastatic index lesions visualized by metabolic imaging. As expected, these samples were enriched for genomic alterations in *TP53*, *RB1*, and *PTEN* or in other alterations linked to these pathways such as *CDKN2A/CDKN2B* (RB1) and *AKT1* (PTEN) (*13*, *33*, *34*) (Fig. 5A). Transcriptomes from 51,094 cells were analyzed (see methods), and the tumor cells ($n = 27{,}338$) revealed notable heterogeneity across samples with a clear distinction based on CRPC-adeno and NEPC histology (Fig. 5B and fig. S15, A to C). Cells were scored for JAK/STAT, FGFR, AR, EMT, and NEPC signatures (see methods), derived from published GEMM and organoid plasticity models (table S18), and a mean value was determined for each PhenoGraph cluster [k-nearest

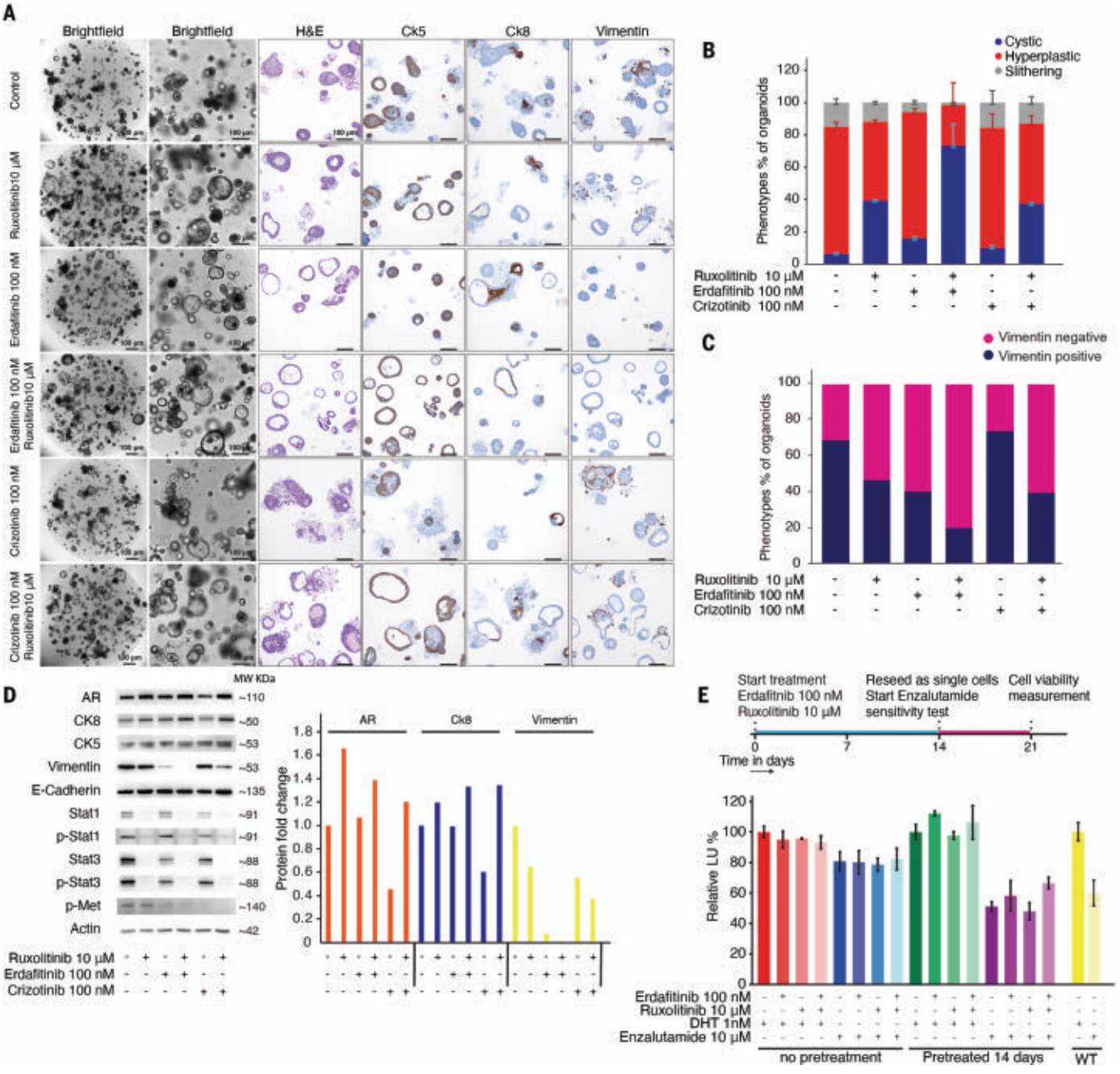


Fig. 4. Pharmacological inhibition of JAK/STAT signaling and FGFR signaling resensitizes prostate cancer organoids to androgen receptor signaling inhibitors. (**A**) Representative brightfield microscopy images, H&E staining, and IHC of CK8, Ck5, and vimentin of *Trp53*$^{\Delta/\Delta}$;*Rb1*$^{\Delta/\Delta}$ organoids treated for 14 days with the indicated drugs. Scale bars represent 100 µm. (**B**) Quantification of phenotypes cystic (blue), hyperplastic (red), and slithering (silver) of *Trp53*$^{\Delta/\Delta}$;*Rb1*$^{\Delta/\Delta}$ organoids treated for 14 days with the indicated drugs; see methods for exact medium composition. (**C**) Quantification of vimentin positivity of organoids in (A). The total number of organoids that were positive for the marker were quantified and normalized. (**D**) On the left is a Western blot of lineage markers and JAK/STAT signaling components in *Trp53*$^{\Delta/\Delta}$;*Rb1*$^{\Delta/\Delta}$ organoids treated for 14 days with the indicated drugs; see methods for exact medium composition. Proteins were probed as indicated. Actin was used as a loading control. Shown on the right is the protein fold change for AR, Ck8, and vimentin as determined by ImageJ analysis in *Trp53*$^{\Delta/\Delta}$;*Rb1*$^{\Delta/\Delta}$ organoids treated for 14 days with the indicated drugs. (**E**) At the top is a schematic overview of the resensitization drug experiments. Organoids are treated with Erda 100 nM and Rux 10 µM for 14 days in low–epidermal growth factor (EGF) organoid medium (ENR+A83+DHT), and control organoids are cultured in low-EGF organoid medium for 14 days. Subsequently, organoids (10,000 cells per well, triplicate) are reseeded in organoid culture medium without EGF (NR+A83) that contains an AR agonist (DHT 1 nM) or antagonist (enzalutamide 10 µM). Viability is shown at the bottom, which was measured by CellTiterGlo after 7 days of enzalutamide treatment of *Trp53*$^{\Delta/\Delta}$;*Rb1*$^{\Delta/\Delta}$ organoids that were treated for 14 days with the indicated drugs. Results were normalized to control-culture conditions. See methods for exact medium composition. The bar graph represents the relative viability to control, as measured by Cell Titer Glo Light Units (LU). Data are presented as mean ± SD.

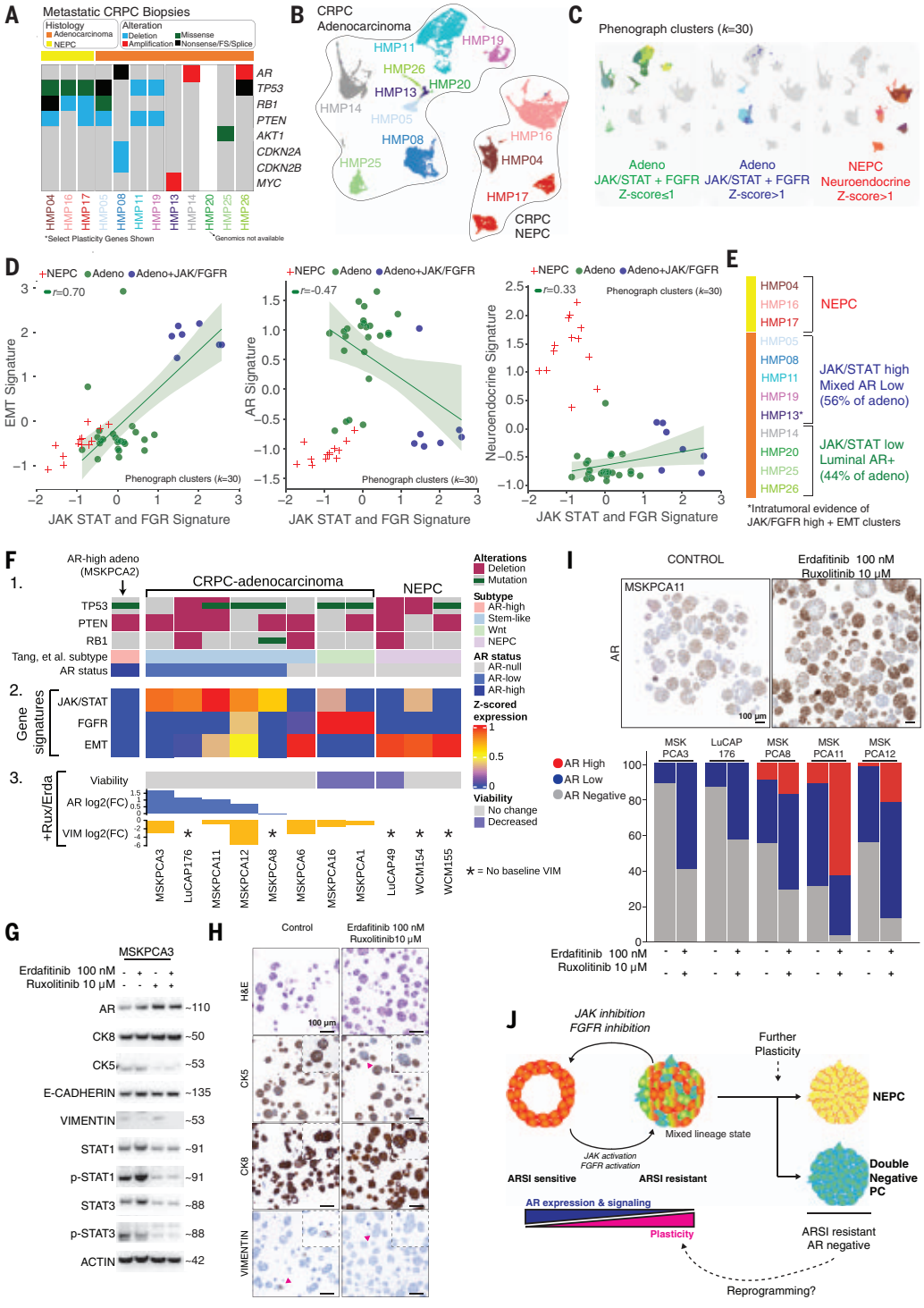


Fig. 5. Activation of JAK/STAT and FGFR signaling arises in CRPC patient biopsies and PDOs. (**A**) Treatment-refractory metastatic CRPC patients underwent interventional radiology–guided biopsies. High-risk index lesions were guided by advanced molecular imaging. Mutational oncoprints were labeled by sample ID (HMP, human metastatic prostate), histology (orange, adenocarcinoma; yellow, neuroendocrine prostate cancer), and mutational status (blue, deletion; red, amplification; green, missense; black, stop or frameshift). (**B**) UMAPs of tumor cells from patient-derived metastatic CRPC (N = 27,338 cells) colored by tumor ID and grouped by tumor type. (**C**) UMAPs of tumor cells from patient-derived metastatic CRPC, annotated by PhenoGraph clusters (k = 30), divided into three groups defined by the Z-score of either JAK/STAT and FGFR or NEPC signatures: (i) AR-high and JAK/STAT+FGFR-low adenocarcinoma (JAK and FGFR signature, Z-score ≤ 1, green), (ii) AR-low and JAK/STAT+FGFR-high adenocarcinoma (JAK and FGFR signature, Z-score > 1, blue), and (iii) NEPC (neuroendocrine signature, Z-score > 1) (see methods). (**D**) Three scatterplots of PhenoGraph clusters (points) based on the Z-score of a combined JAK/STAT+FGFR signature (x axis) versus EMT, AR, and NEPC signatures (y axis), respectively. Linear fits were calculated for adenocarcinoma clusters only; the corresponding Pearson's correlation is noted. Clusters annotated by groups defined in (C): NEPC (red), AR-high and JAK/FGFR-low adeno (green), and AR-low and JAK/FGFR-high adeno (blue). (**E**) Proposed classification of metastatic CRPC samples based on JAK/STAT+FGFR and NEPC

signatures. HMP13 is assigned to the JAK/STAT+FGFR-high group, because this sample harbors a well-demarcated JAK/FGFR and EMT-high subpopulation, described in fig. S14D. (**F**) Schema of baseline attributes and functional change in human organoids with null to low AR expression after treatment with combined Rux and Erda, with hold-out sample MSKPCA2 used as reference level for the AR-high subtype. Shown are (1) baseline attributes, including an oncoprint of the *TP53*, *RB1*, and *PTEN* genotypes, transcriptional subtypes based on Tang *et al*. (*36*), and low versus null AR protein expression; (2) a heatmap of baseline gene signatures in the human organoids using publicly available bulk RNA-seq based on mean *Z*-scored expression of *JAK/STAT* signaling, *FGFR* signaling, and *EMT* gene sets (see methods); and (3) functional changes in human organoids with null to low *AR* expression after treatment with Rux and Erda, including cell viability, and log2 fold change (FC) in AR and VIM protein expression. (**G**) Western blot of lineage markers and JAK/STAT signaling components in MSKPCA3 organoid after 14 days of treatment with Erda 100 nM, Rux 10 μM, or with Erda 100 nM and Rux 10 μM combination. Proteins were probed as indicated. Actin was used as a loading control. (**H**) Representative histology and IHC (CK8, CK5, and vimentin) of MSKPCA3 organoids after 14 days of treatment with Erda (100 nM) and Rux (10 μM) combination in full organoid medium (ENRFFPN, A83-01, nicotinamide, DHT). Red arrowheads indicate organoids of interest with changed protein expression. See methods for exact medium composition. Scale bars represent 100 μm. (**I**) Shown at the top is a representative staining of AR in control-treated MSKPCA11 organoids and those treated with Erda 100 nM and Rux 10 μM. Shown at the bottom is a bar graph of AR IHC staining patterns in PDOs MSKPCA3, LuCAP176, MSKPCA8, MSKPCA11, and MSKPCA12. Organoids were control-treated or treated with Erda 100 nM and Rux 10 μM. AR staining was classified as negative (gray), low intensity (blue), and high intensity (red). (**J**) Proposed model system of lineage plasticity. JAK/STAT signaling activation and FGFR signaling activation leads to an AR-low, androgen receptor signaling inhibitor (ARSI)–insensitive state that can be reprogrammed back to an ARSI-sensitive state. Potentially, the JAK/STAT-high, AR-low, ARSI-insensitive state is a cellular state preceding AR^-, ARSI-insensitive CRPC. Mechanisms leading to AR^- CRPC are unknown.

neighbors (k) = 30]. Scatterplots of these clusters revealed a strong positive correlation between JAK/STAT with FGFR signaling and EMT [correlation coefficient (r) = 0.70], but a negative correlation with AR signaling (r = −0.47) (Fig. 5D). One particular CRPC adenocarcinoma tumor (HMP13) distinctively displayed intratumoral heterogeneity, harboring a subpopulation marked by JAK/STAT and FGFR up-regulation, EMT, and down-regulation of AR signaling (fig. S15D and table S20). Mapping of TFs in this sample across pseudotime (*25*) revealed a similar activation of transcriptional regulators, namely STAT2/3, IRF1, and IRF7 (fig. S15, E and F, for robustness analysis). Overall, clusters with JAK/STAT with FGFR- and EMT-high programs are found in 5 out of 9 (or 56%) of CRPC adenocarcinoma samples (Fig. 5E; defined by having at least >20 cells in a cluster). These results are highly consistent with our findings from the GEMM and organoid models where JAK/STAT and FGFR are strongly correlated with EMT programs and lower AR signaling but absent in NEPC cells.

To address whether the JAK/STAT and FGFR activation we observed in human CRPC samples is functionally linked to lineage plasticity phenotypes such as EMT and NEPC, we turned to a panel of human prostate cancer organoids derived from patients with CRPC, all of which have been characterized by whole-exome sequencing and bulk RNA-seq and represent the CRPC-adeno and NEPC histology phenotypes described earlier. As expected, these patient-derived organoids (PDOs) were enriched for alterations in *TP53*, *RB1*, and *PTEN*. For the CRPC-adeno subgroup, we have previously shown that those with high AR expression (e.g., MSKPCa2) express luminal lineage markers and remain sensitive to Enz (*35*). We therefore focused on CRPC-adeno organoids with low or absent AR expression, as well as PDOs with NEPC histology, to determine if Rux with Erda treatment might restore expression of luminal genes as we saw in *Trp53* and *Rb1*–deficient mouse organoids. Transcriptomic analysis revealed increased JAK/STAT activation and EMT signatures in five PDOs (MSKPCa3, 8, 11, and 12 and LuCap176), all of which have low AR expression and therefore reflect the JAK/STAT-high, mixed AR-low subgroup seen in the patient biopsies (Fig. 5F, fig. S16A, and table S21). Unlike our genetically defined mouse organoid model, the human PDOs contain a complex array of genomic alterations beyond the tumor suppressor gene annotated in Fig. 5F and have growth properties that, in general, are much slower (4- to 7-day doubling times) than mouse organoids (1 to 2 days). In addition, individual PDOs often have heterogeneous expression of lineage markers, exemplified by variations in the level and percentage of cells with AR expression when examined on a cell-by-cell basis (Fig. 5, G to I).

Despite these differences between the mouse and human organoid models, we observed increased AR expression after 14 days of Rux with Erda treatment in four of the five AR-low and JAK/STAT-high PDOs. This change in AR expression was accompanied by increased expression of the luminal cytokeratin CK8 and decreased expression of the mesenchymal marker (VIM) (Fig. 5, F to I, and figs. S16B and S17). As with the *Trp53* and *Rb1*–deleted mouse organoid, Rux with Erda treatment did not affect the proliferation rate of the AR-low and JAK/STAT-high PDOs (fig. S18, A and B). One important distinction from the mouse system, however, is that a substantial percentage of cells within each of these Rux-with-Erda–sensitive organoids remained AR^- (Fig. 5I). Whether this is a consequence of the greater genomic complexity of the human versus mouse models and/or the need for longer periods of drug treatment to reprogram human organoids fully (owing to their slower growth rate) requires further study.

In addition to the luminal reprogramming phenotype seen in the AR-low, JAK/STAT-high organoids, we noted a distinct phenotype of substantial growth inhibition in two AR^- PDOs with CRPC-adeno histology (MSKPCa1 and 16) (fig. S18, A and B). These PDOs are noteworthy for signatures of highly activated FGFR signaling but modest or absent JAK/STAT activation and likely reflect a dependency on FGFR for proliferation. Indeed, evidence of FGFR-dependent CRPC has been previously reported in the context of AR^-, SYP^- (double negative) prostate cancer (*4*). Growth of one of the three NEPC PDOs (LuCaP49) was also notably inhibited by Rux with Erda despite low FGFR and JAK/STAT activation states across all three lines. The mechanism underlying this sensitivity requires further investigation.

Finally, we mapped the Rux-with-Erda sensitivity profiles to four subtypes of human CRPC (AR^+, Stem-like, Wnt, and NEPC) recently reported based on integrated chromatin accessibility and transcriptome analysis (*36*). The four PDOs that display luminal reprogramming all fall within the stem-like subtype [also defined by increased JAK/STAT activation in Tang *et al*. (*36*)], whereas the two CRPC-adeno PDOs with growth inhibition fall with the Wnt subtype (Fig. 5F), providing evidence of specific and exploitable therapeutic vulnerabilities for molecularly distinct subtypes of human CRPC.

Discussion

Across multiple cancers, marked changes in tumor cell state are implicated in the resistance to targeted therapies. This lineage plasticity is particularly relevant in the context of next-generation drugs designed to overcome conventional resistance that arises from drug-target mutations. However, lineage transformation remains poorly characterized, and we even lack a basic quantitative description of phenotypic plasticity. To address this mechanism of resistance therapeutically, it is critical to understand when, where, and how cell-state changes arise—ideally, by profiling tumor cells as resistance arises. We performed scRNA-seq time-course experiments on prostate cancer GEMMs and organoids, enabling an in-depth study of the adenocarcinoma-to-NEPC transition upon deletion of tumor suppressor gene

combinations relevant to human CRPC (*Trp53*, *Rb1*, *Pten*). Our analysis identified an adenocarcinoma population with mixed luminal and basal identity that emerges within 2 weeks of tumor suppressor loss and progresses to EMT or NEPC phenotypes. We developed approaches to quantify plasticity, showing that it arises in a cell-autonomous fashion in both normal luminal and basal cells and increases as migratory behavior and EMT-like transcriptional signatures are acquired. However, progression to an NEPC phenotype only occurs in vivo, presumably reflecting a need for stromal or immune cell signals that are missing from organoid culture.

Prior work (*4*, *15*, *16*, *28*) established that the onset of lineage plasticity is accompanied by resistance to AR pathway inhibition, which we hypothesize to be a consequence of acquiring previously uncharacterized AR-independent cell states, including basal or mesenchymal lineages. Importantly, we also observed the converse—plasticity is initially driven by loss of tumor suppressor genes, and subsequently accelerated by AR pathway inhibition. The fact that lineage plasticity, which is initially responsible for drug resistance, is exacerbated by further drug treatment has important medical implications.

We searched for molecular drivers of plasticity by identifying therapy-induced gene programs associated with early acquisition of our newly quantifiable high-plasticity state and discovered an unexpected, robust signature of JAK/STAT activation in both organoids and GEMMs. Other manifestations of plasticity such as EMT and NEPC likely evolve later, based on the temporal sequence observed in organoids as well as absence of JAK/STAT activation in the NEPC component of GEMM tumors and human NEPC samples. This observation suggests an early "critical period" after the onset of plasticity, during which therapies might prevent, delay, or even reverse plasticity. Supporting this hypothesis, JAK (and FGFR) inhibitors restored expression of luminal markers and decreased mesenchymal markers in mouse organoids and human PDOs, but only in those with increased JAK/STAT activation. These changes evolved over several weeks of treatment in the absence of growth inhibition, consistent with a lineage reprogramming mechanism akin to differentiation therapy used in subclasses of acute myeloid leukemia (*37*, *38*).

By searching for autocrine loops associated with JAK/STAT activation, we discovered that FGFR signaling increases with the onset of plasticity in organoids, GEMMs, and human CRPC samples. Using a highly selective pan-FGFR inhibitor (Erda), we found (i) potent luminal lineage reprogramming in combination with JAK inhibition (Rux) in mouse organoids and PDOs, perhaps due to cross-talk between these pathways and reprogramming transcription factors or chromatin modifiers; and (ii) consistently reduced mesenchymal gene (VIM) expression in organoids after FGFR inhibition, independent of JAK/STAT activation. This is consistent with FGFR's role in tumors with an EMT phenotype, as proposed in AR^-, SYP^- prostate cancer and in breast cancer (*39*). Importantly, the primary consequence of FGFR inhibition in PDOs with low JAK/STAT activation is growth inhibition, suggesting a dependency on FGFR for survival, which is distinct from the lineage reprogramming phenotype in PDOs with high JAK/STAT activation. Collectively, these findings suggest a therapeutic strategy by which CRPC patients with JAK/STAT-high, AR-low signaling [defined as stem-like by Tang *et al.* (*36*)] would be candidates for combined JAK and FGFR inhibition in order to reprogram tumor cells to an AR-high luminal state and restore sensitivity to AR-targeted therapy. Those with increased FGFR signaling and no JAK/STAT activation [defined as Wnt subtype by Tang *et al.* (*36*)] would also be candidates for FGFR inhibition to exploit FGFR dependency for survival. Therapeutic timing is critical, as shown by cell-state heterogeneity within a single CRPC patient (or GEMM tumor) and absence of JAK/STAT activation in NEPC. These findings provide a rationale for molecular diagnostic assays that could aid patient selection.

REFERENCES AND NOTES

1. L. V. Sequist *et al.*, *Sci. Transl. Med.* **3**, 75ra26 (2011).
2. M. M. Awad *et al.*, *N. Engl. J. Med.* **384**, 2382–2393 (2021).
3. P. A. Watson, V. K. Arora, C. L. Sawyers, *Nat. Rev. Cancer* **15**, 701–711 (2015).
4. E. G. Bluemn *et al.*, *Cancer Cell* **32**, 474–489.e6 (2017).
5. Á. Quintanal-Villalonga *et al.*, *Nat. Rev. Clin. Oncol.* **17**, 360–371 (2020).
6. A. J. Schoenfeld *et al.*, *Clin. Cancer Res.* **26**, 2654–2663 (2020).
7. I. Tirosh *et al.*, *Science* **352**, 189–196 (2016).
8. D. J. Konieczkowski *et al.*, *Cancer Discov.* **4**, 816–827 (2014).
9. F. Rambow *et al.*, *Cell* **174**, 843–855.e19 (2018).
10. A. M. Laughney *et al.*, *Nat. Med.* **26**, 259–269 (2020).
11. M. X. He *et al.*, *Nat. Med.* **27**, 426–433 (2021).
12. B. Dong *et al.*, *Commun. Biol.* **3**, 778 (2020).
13. D. Robinson *et al.*, *Cell* **161**, 1215–1228 (2015).
14. S. G. Zhao *et al.*, *JAMA Oncol.* **3**, 1663–1672 (2017).
15. H. Beltran *et al.*, *Cancer Discov.* **1**, 487–495 (2011).
16. S. Y. Ku *et al.*, *Science* **355**, 78–83 (2017).
17. Y.-H. Huang *et al.*, *Genes Dev.* **32**, 915–928 (2018).
18. N. J. Brady *et al.*, *Nat. Commun.* **12**, 3372 (2021).
19. R. C. Fitzgerald *et al.*, *Lancet* **396**, 333–344 (2020).
20. W. R. Karthaus *et al.*, *Science* **368**, 497–505 (2020).
21. L. Crowley *et al.*, *eLife* **9**, e59465 (2020).
22. W. Guo *et al.*, *Nat. Genet.* **52**, 908–918 (2020).
23. S. Nowotschin *et al.*, *Nature* **569**, 361–367 (2019).
24. J. Cao *et al.*, *Nature* **566**, 496–502 (2019).
25. M. Setty *et al.*, *Nat. Biotechnol.* **37**, 451–460 (2019).
26. W. R. Karthaus *et al.*, *Cell* **159**, 163–175 (2014).
27. C. S. Kuo, M. A. Krasnow, *Cell* **163**, 394–405 (2015).
28. P. Mu *et al.*, *Science* **355**, 84–88 (2017).
29. Z. Zhou *et al.*, *Cancer Res.* **66**, 7889–7898 (2006).
30. A. Davies *et al.*, *Nat. Cell Biol.* **23**, 1023–1034 (2021).
31. S. Persad *et al.*, bioRxiv 2022.04.02.486748 [Preprint] (2022); https://doi.org/10.1101/2022.04.02.486748.
32. A. Decout, J. D. Katz, S. Venkatraman, A. Ablasser, *Nat. Rev. Immunol.* **21**, 548–569 (2021).
33. H. Beltran *et al.*, *Nat. Med.* **22**, 298–305 (2016).
34. W. Abida *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11428–11436 (2019).
35. D. Gao *et al.*, *Cell* **159**, 176–187 (2014).
36. F. Tang *et al.*, *Science* **376**, eabe1505 (2022).
37. R. J. Stubbins, A. Karsan, *Blood Cancer J.* **11**, 193 (2021).
38. H. de Thé, *Nat. Rev. Cancer* **18**, 117–127 (2018).
39. D. Georgopoulou *et al.*, *Nat. Commun.* **12**, 1998 (2021).

ACKNOWLEDGMENTS

We thank the Pe'er lab and the Sawyers lab for valuable critiques and discussions. We are grateful to the prostate cancer patients who participated in this research. We further appreciate the efforts of the Memorial Sloan Kettering Cancer Center (MSKCC) Genitourinary faculty for recruitment of prostate cancer tumor specimens, the Molecular Cytology Core Facility from MSKCC for help with confocal microscopy and IHC, and the Flow Cytometry Core Facility from MSKCC for help with FACS experiments.

Funding: This work is supported by the Howard Hughes Medical Institute (HHMI) (C.L.S. and D.P.); National Institute of Health grants CA193837 (C.L.S.), CA092629 (C.L.S. and Y.C.), CA224079 (C.L.S. and Y.C.), CA155169 (C.L.S.), CA008748 (C.L.S., D.P., and Y.C.), and CA209975 (D.P.); the Starr Cancer Consortium (I12–0007); the Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center (D.P., J.M.C., O.C., I.M., and L.M.); NIH K08 CA259161-01A1 (J.M.C.); an American Association for Cancer Research Lung Cancer Fellowship (J.M.C.); American Society of Clinical Oncology Young Investigator Awards (J.M.C., S.Z., and J.L.Z.); a Department of Defense Prostate Cancer Research Program Early Investigator Research Award (S.Z.); Prostate Cancer Foundation PCF 20YOUN10 (S.Z); a Prostate Cancer Foundation Young Investigator Award (J.L.Z.); the Louis V. Gerstner, Jr. Physician Scholars Program (S.Z.); a Career Development Award in Clinical Oncology from the National Cancer Institute (J.L.Z.); a Movember GAP2 grant and a Prostate Cancer Foundation Challenge Award (M.J.M.); and Prostate Cancer Foundation PCF 17YOUN10 (W.R.K.). **Author contributions:** W.R.K. and C.L.S. designed organoid experiments. W.R.K. and J.R.L. performed organoid experiments and downstream benchwork. J.M.C., D.P., M.S., Z.-N.C., and S.P. performed computational method development. J.M.C., S.Z., and M.S. performed computational analysis. J.C. assisted with computational analyses. J.R.L., J.L.Z., J.L., K.E.L., and W.R.K. performed mouse work. S.Z., W.R.K., and M.J.M. obtained and processed human tissue. I.M., O.C., and T.X. performed single-cell experiments. A.B. performed IHC and IF, I.L. performed IHC on human tissues, and A.G. performed pathology on human samples. D.W.G. provided GEMMs. S.W., C.L., and Y.C. provided PDOs. L.M. and R.C. oversaw single-cell experiments. W.R.K., D.P., and C.L.S. conceived the project. D.P., C.L.S., J.M.C., W.R.K., and S.Z. wrote the manuscript. D.P. and C.L.S. oversaw the project. **Competing interests:** C.L.S. is on the board of directors of Novartis, is a cofounder of ORIC Pharmaceuticals, and is a co-inventor of the prostate cancer drugs enzalutamide and apalutamide, covered by US patents 7,709,517; 8,183,274; 9,126,941; 8,445,507; 8,802,689; and 9,388,159 filed by the University of California. C.L.S. is on the scientific advisory boards of the following biotechnology companies: Agios, Beigene, Blueprint, Column Group, Foghorn, Housey Pharma, Nextech, KSQ Therapeutics, Petra Pharma, and PMV Pharma and is a cofounder of Seragon Pharmaceuticals, purchased by Genentech/Roche in 2014. D.P. is on the scientific advisory board of Insitro. W.R.K. is a co-inventor of organoid technology. J.L.Z. is currently a full-time employee of AstraZeneca. AstraZeneca plays no role in the current study. Y.C. has stock ownership in ORIC Pharmaceuticals and receives research funding from Foghorn Therapeutics. **Data and materials availability:** Sequencing data are deposited at the Gene Expression Omnibus (GEO) database (GSE210358). Code for scBLender and Jupyter notebooks to reproduce figures are available for download at https://github.com/dpeerlab/Lineage_Plasticity_in_CRPC. All other data are available in the manuscript or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn0478
Materials and Methods
Figs. S1 to S18
Tables S1 to S24
References (*40*–*80*)
MDAR Reproducibility Checklist
Movies S1 and S2

Submitted 30 October 2021; resubmitted 28 April 2022
Accepted 5 August 2022
Published online 18 August 2022
10.1126/science.abn0478

CORRELATED MATERIALS

Systematic electronic structure in the cuprate parent state from quantum many-body simulations

Zhi-Hao Cui, Huanchen Zhai, Xing Zhang, Garnet Kin-Lic Chan*

The quantitative description of correlated electron materials remains a modern computational challenge. We demonstrate a numerical strategy to simulate correlated materials at the fully ab initio level beyond the solution of effective low-energy models and apply it to gain a detailed microscopic understanding across a family of cuprate superconducting materials in their parent undoped states. We uncover microscopic trends in the electron correlations and reveal the link between the material composition and magnetic energy scales through a many-body picture of excitation processes involving the buffer layers. Our work illustrates a path toward a quantitative and reliable understanding of more complex states of correlated materials at the ab initio many-body level.

In spite of our qualitative theoretical understanding of many electronic phases of matter, there remains a deficit in the quantitative understanding of correlated electron materials (*1*, *2*). This limits our ability to connect the atomic structure and composition to the electronic phenomena, as well to answer fundamental physical questions related to microscopic mechanisms.

Among correlated quantum materials, the high-temperature (T_c) superconductors remain a fertile source of new physics (*3*–*6*). We focus on the cuprates, where one finds the highest superconducting T_c in the mercury-barium cuprate family (*7*). Although progress has been made in understanding the universal phase diagram through numerical calculations on lattice models, the understanding of properties of individual compounds remains largely empirical, with substantial difficulties in linking the observed trends to model parameters.

In principle, a quantitative understanding is simply a matter of many-electron quantum mechanics, but solving the Schrödinger equation beyond lattice models involves three challenges: the quantum many-body correlations, the thermodynamic limit (TDL), and the high-energy degrees of freedom and long-range interactions of real materials. We adopt a pragmatic computational framework in which the challenges can be tackled simultaneously: solvers for the many-body problem beyond models (*8*, *9*), self-consistent quantum embedding to develop phases in the TDL (*10*–*12*), and periodic quantum chemistry using local bases (*13*, *14*) to efficiently treat long-range interactions and high-energy degrees of freedom. Each component has been individually tested in prior work, but the principal feature of our combined strategy is that the solution process bypasses models with uncontrolled parameters; the only remaining parameters are the size of the computational cell, the basis size, and the level of the many-body solver. Thus, all aspects of the calculation can in principle be controlled toward exactness.

In this work, we describe the full application of this strategy to the simulation of a family of cuprates in their parent phase at zero temperature. Although the parent phase is qualitatively simple, and elements of our framework have been used to understand exotic physics in simplified models (*15*), obtaining quantitative material systematics and functional relationships even in the parent phase is a major challenge and serves as a litmus test of the promise of our overall strategy. Our detailed simulations improve the level of resolution of the electronic structure and enable us to uncover direct links between the material-specific physics and composition.

Cuprates and the parent state

Structure

The main structural feature of the cuprates is the two-dimensional (2D) CuO_2 (formally $[CuO_2]^{2-}$) square lattice plane (Fig. 1A). In different cuprates, the copper-oxygen plane is surrounded by different atoms and buffer layers in the vertical direction. We consider three specific compounds, in addition to layer-stacked idealized CuO_2 planes (geometries are provided in table S1). The first is infinite layer $CaCuO_2$ (CCO) (Fig. 1D), where calcium counterions intercalate between the CuO_2 planes in an infinitely repeating structure. CCO does not itself superconduct because of difficulties in doping the material. However, high T_c are observed in the related mercury-barium cuprates (the Hg–Ba–Ca–Cu–O family). There, the CuO_2 plane is decorated by apical oxygens, which connect to buffers of Hg and Ba ions. Unlike in CCO, the buffer layers form large spacers between the copper-oxygen layers. Different mercury-barium cuprates can be synthesized with different numbers of CuO_2 planes between each buffer layer, leading to single-layer cuprates, double-layer cuprates, and so on. We consider two members in this family: $HgBa_2CuO_4$ (Hg-1201; single-layer, T_c = 97 K) and $HgBa_2CaCu_2O_6$ (Hg-1212; double-layer, T_c = 127 K). Hg-1201 exhibits distorted octahedral Cu–O coordination (two apical oxygens per Cu) (Fig. 1B), whereas each layer of Hg-1212 contains pyramidal Cu–O coordination (one apical oxygen per Cu) (Fig. 1C). Hg-1201, Hg-1212, and CCO are compositionally related by replacing Hg-Ba-apical O layers with Ca layers.

Parent state

Unlike conventional superconductors, the parent state of the cuprates is an antiferromagnetic (AFM) insulator with long-range order, resulting from the strong Cu d – d electron interaction. Typical Néel temperatures for the AFM state range from about 250 K (in Nd_2CuO_4) to 450 K (in $YBa_2Cu_3O_6$) (*5*), and only after doping does the ground state enter the superconducting phase. It is generally thought that the antiferromagnetism is to first order approximated by 2D nearest-neighbor (NN) Heisenberg-like physics. However, the 2D NN Heisenberg model does not reproduce the dispersion of the experimental spin-wave spectrum, and questions remain as to the magnitude, sign, and material-specific origin of corrections to the NN picture.

There have been many attempts to correlate properties of the cuprates in the superconducting phase (such as T_c) with structure, composition, and band structure (*16*–*21*). However, without a direct ability to simulate the material T_c with different parameters, distinguishing correlation from causation is difficult. Although there has been less focus on correlating properties of the parent state with physical features, many proposals relate the high Néel temperatures and strong exchange coupling in the parent state to the superconducting mechanism and other exotic physics under doping. We establish causal, quantitative relationships between the magnetic features of the parent state and the atomic-scale structural and electronic features of the materials.

Theoretical techniques

Strategy

Previous numerical work on cuprate electronic structure [with a few exceptions, such as (*22*, *23*)] falls in two classes: (i) ab initio all-electron simulations with a modest treatment of electron correlations (*24*–*26*), which are often used to derive low-energy effective models, and (ii) accurate many-body methods applied to low-energy effective models, to obtain phase diagrams and more exotic orders (*27*–*33*).

Our strategy is to use families of methods associated with the model studies of (ii), but technologically elevated to the fully ab initio

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.
*Corresponding author. Email: gkc1000@gmail.com

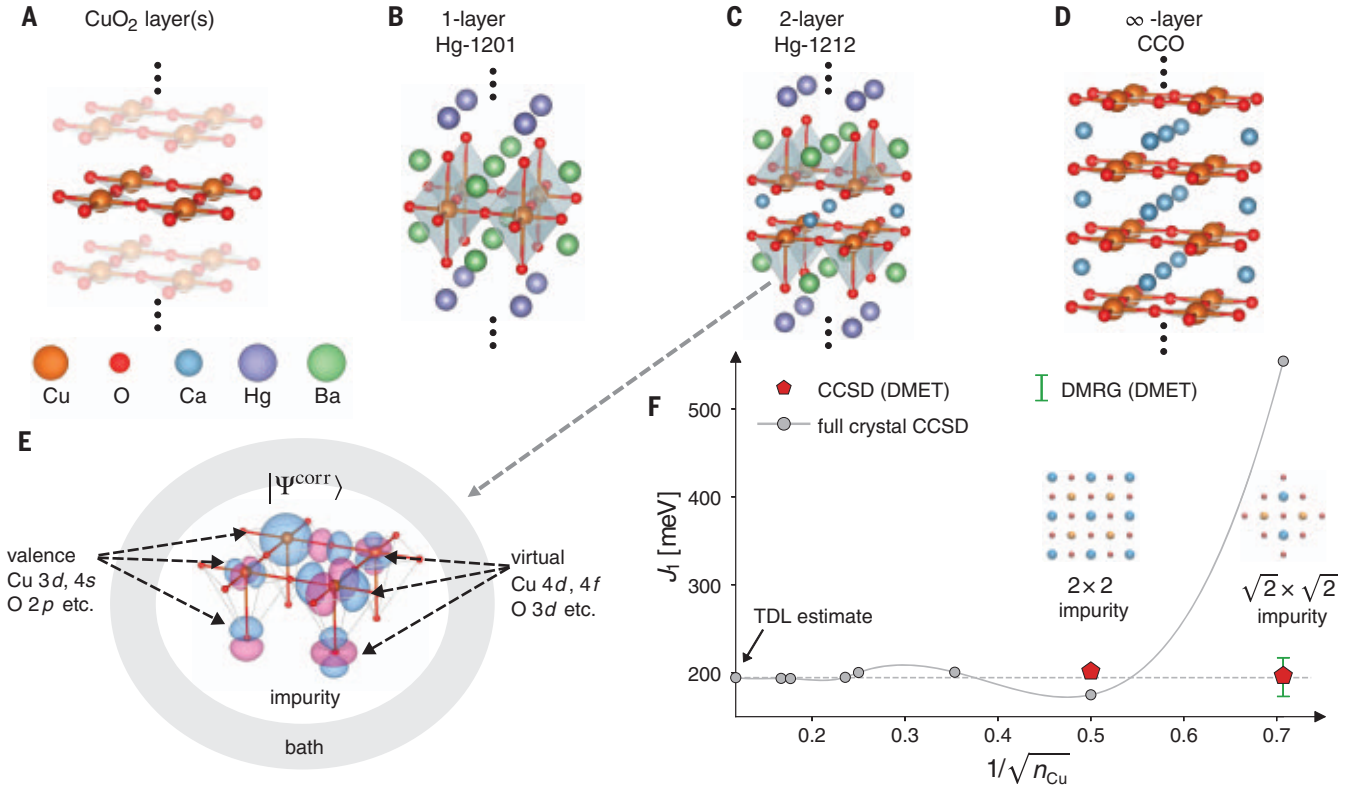


Fig. 1. Structures and computational strategy. (**A**) $[CuO_2]^{2-}$ plane(s) in cuprates. (**B** to **D**) Relationship between single-layer Hg-1201 ($HgBa_2CuO_4$), double-layer Hg-1212 ($HgBa_2CaCu_2O_6$), and infinite layer CCO ($CaCuO_2$). Ca layers replace the Hg-Ba-apical-O layers. (**E**) The ab initio density matrix embedding framework. The Hg-1212 lattice is divided into an impurity (for example, the 2 × 2 cell) with the environment replaced by a bath; the atoms are represented by local valence and virtual orbitals, and the impurity problem is solved for the many-body wave function Ψ^{corr}. (**F**) Correlation and finite size effects in the NN Heisenberg exchange coupling J_1. We compare the exchange coupling (*y* axis) from a full-crystal CCSD calculation as a function of CCO crystal size [plane side-length in units of Cu atoms, n_{Cu} (*x* axis)] in a small basis, to embedded calculations with two impurity sizes and two solvers, DMRG and CCSD. The embedded 2 × 2 impurity is already close to the TDL, and the DMRG and CCSD impurity solvers agree well in the smallest impurity.

Hamiltonians of (i). This bypasses the ambiguities of intermediate downfolded models while allowing correlated physics to emerge. The three numerical components are the quantum embedding, the ab initio all-electron infrastructure, and the many-body solvers. Our technical setup uses density matrix embedding theory (DMET) to self-consistently embed a 2 × 2 supercell (impurity) of the cuprate material within an all-electron description, and we solve the resulting embedded impurity with an ab initio many-body approximation [coupled cluster (CC) theory]. To do so feasibly and reliably relies on recent advances and techniques introduced in this work, such as a subimpurity formalism and improved DMET self-consistency algorithms for large impurities, improved ab initio matrix element generation, and careful solver benchmarking against a massively parallel ab initio density matrix renormalization group (DMRG) implementation. Below, we describe the quantum embedding and many body solvers; the ab initio infrastructure is discussed in (*34*), section 1.1.

The quantum embedding provides a framework for phases that emerge as a result of interactions (*35*) and includes dynamical mean-field theory and its relations (*36*–*38*) and the DMET (*39*, *40*) used in this work. We separate the material into an impurity region and a bath that describes fluctuations out of the impurity, and their self-consistency yields emergent phases. The embedding becomes exact with increasing impurity size.

In previous work on the one-band and three-band Hubbard models, DMET has been extensively benchmarked against other methods and, for example, accurately resolves exotic order in the underdoped region (*15*, *41*). (The ability of DMET to capture exotic physics in doped lattices suggests a path from the ab initio studies of the parent state here to the physics of the doped materials.) To move beyond models to the ab initio physics, we start from our recently introduced all-electron, full-cell approach (*10*–*12*). In this work, the impurity is a supercell of the cuprate that contains all atoms and orbitals, with all quartic interactions between the orbitals. In contrast to downfolded approaches with a handful of impurity orbitals and possibly simplified interactions (*37*, *42*), our largest impurity (in Hg-1212) contains 48 atoms and close to 900 orbitals (Fig. 1E). These orbitals include many "virtual" bands, which capture quantitative electron correlation effects and screening. Part of the reason why these large impurities are feasible is the DMET formulation itself, which bypasses expensive frequency-dependent quantities. The other critical factors are the choice of solvers and the periodic quantum chemistry infrastructure based on local atomic basis sets, which compactly discretize the virtual bands for electron correlation [(*34*), section 1].

The quantum impurity problem in the full-cell approach is a many-body problem with hundreds of orbitals. This can be solved because many orbitals do not display strongly correlated physics. We use two impurity solvers in this work. We obtain the majority of the results using an ab initio CC singles and doubles (CCSD) (*43*) solver. Although approximate, they treat clusters of (arbitrarily) strongly correlated particles exactly and have previously been shown to yield accurate results in various quantum impurity problems (*9*–*12*, *44*), particularly in ordered phases. To

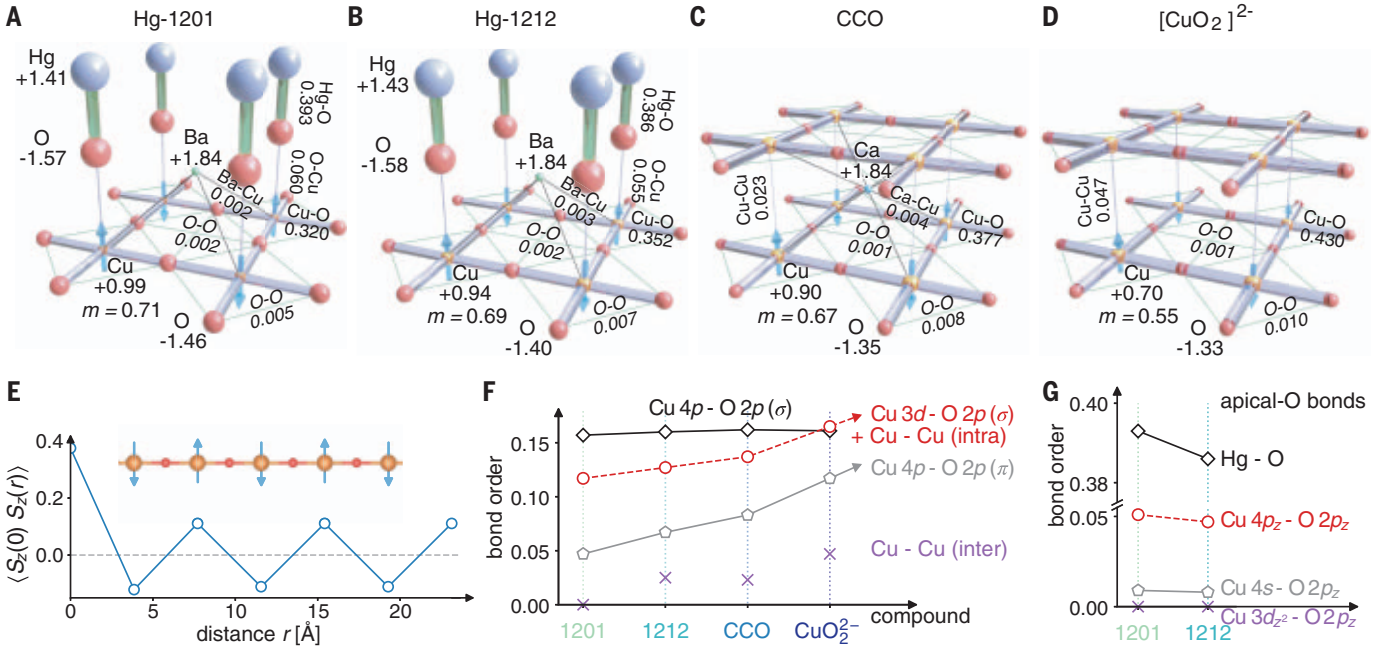


Fig. 2. Charge, spin, and bond orders. (**A** to **D**) Charge *c*, magnetic moment *m*, and bond order *b* of different cuprates. Cu, yellow; O, red; Hg, violet; Ba, green; Ca, blue. Atomic sphere radius is the number of electrons *n* (local charge $Z - n$ is labeled; *Z*, nuclear charge); arrow length indicates the magnitude of local moment $m = n^{\uparrow} - n^{\downarrow}$; and bond width indicates the bond order *b*. (**E**) Spin-spin correlation function $\langle S_z(0)S_z(r)\rangle$ in CCO. (**F**) Cu orbital-resolved bond orders for various compounds. (**G**) Apical O orbital-resolved bond orders for Hg-1201 and Hg-1212. More details are available in (*34*), section 2.2.

verify the accuracy of the CC approximation, we use a second solver, the quantum chemistry DMRG (*8*, *45*), to benchmark a subset of problems.

Computational setup

The 2 × 2 supercell impurities are shown in Fig. 1 for the different mercury-barium cuprates. In (*34*), section 2.1.5, we also discuss a benchmark study of lanthanum copper oxide. Every atom is represented in a valence double-ζ with polarization basis (def2-SVP) (*46*); for example, each Cu is represented by [*5s3p2d1f*] shells and each O by [*3s2p1d*] shells, and the embedding lattice is chosen to be an 8 × 8 × 2 lattice of the primitive cell. Large impurities (such as in Hg-1212) were further fragmented into smaller subimpurities with up to 364 orbitals (280 impurity orbitals and 84 valence bath orbitals), and impurity solutions were obtained by using CCSD or DMRG. [Unless otherwise indicated, data are from the CCSD solver; DMRG data are in (*34*), section 2.1.4.] The DMET equations were then solved with self-consistency and valence-shell lattice-impurity density matrix fitting.

Benchmarks

Within the above strategy, the only sources of error are from the finite size of the impurity (and embedding lattice), the approximate nature of the impurity solver, and the finite size of the local atomic basis. We carried out extensive benchmarking to verify the specific approximations. In Fig. 1F, we compare results from finite impurities to the TDL (which can be estimated from a full crystal calculation within a small local atomic basis) for the energy difference between the ferromagnetic (FM) and the AFM state (proportional to the NN exchange coupling J_1). We also show the deviation between this energy difference estimate from the DMRG and CCSD solvers in a small impurity where DMRG is tractable. Both sets of data illustrate that the TDL and many-body character of the physics are well captured within the approximations in this work. Additional benchmarks (such as basis set convergence) can be found in (*34*), sections 2.1 and 3.3 (figs. S4 and S5 and tables S15 to S18).

Multiorbital electronic structure

We start with general electronic trends across the series Hg-1201, Hg-1212, CCO, and $[CuO_2]^{2-}$ as a baseline to understand trends in the physics in later sections.

Order parameters and bonding

We first extract order parameters from the 2 × 2 computational supercell: charge, local moment, bond orders (from the off-diagonal elements of $\gamma_{ij} = \langle a_j^{\dagger} a_i \rangle$, where *i*, *j* label local atomic orbitals in the cell), and the spin correlation function $\langle S_z(0)S_z(r)\rangle$ measured across the full crystal (Fig. 2, A to E).

The key features are the following: (i) The ground state is AFM with long-range order, with the moment in the Cu half-filled $3d_{x^2-y^2}$ orbital. Cu *4s*/*4p* occupancy reduces the total moment by about 10%. The unit cell moment ranges from 0.71 in Hg-1201 to 0.55 in $[CuO_2]^{2-}$. (ii) Charge is transferred from in-plane O orbitals to the other ions, with the degree of transfer increasing across the series. There is significant charge transfer to the Cu minority spin orbitals (as much as 0.3 electrons in $[CuO_2]^{2-}$). (iii) Ca and Ba buffer atoms in CCO, Hg-1201, and Hg-1212 are ionic, with Hg covalently bonded to the apical oxygen through the O $2p_z$-Hg *6s*, $5d_{z^2}$ bonds. Hg-1201 and Hg-1212 do not differ much with respect to the out-of-plane observables but do differ for their observables in the CuO_2 plane. (iv) In-plane σ-bonding (Fig. 2F) is predominantly Cu *4p*-O *2p* and does not differ much across the compounds. However, Cu *3d*-O *2p* bonding and out-of-plane π bonding increase across the series, reflecting increasing in-plane *3d*/*4p* hybridization. The change in bonding is not (solely) caused by the structural changes (for example, CCO and $[CuO_2]^{2-}$ have the same Cu–O bond length but different bond orders) but instead reflect redistribution of charge from the buffer layers. (v) The apical oxygen bond order (Fig. 2G) decreases from Hg-1201 to Hg-1212, with the oxygen only weakly bound to Cu. Cu *4s* and $4p_z$ contribute to apical bonding, with little $3d_{z^2}$ participation.

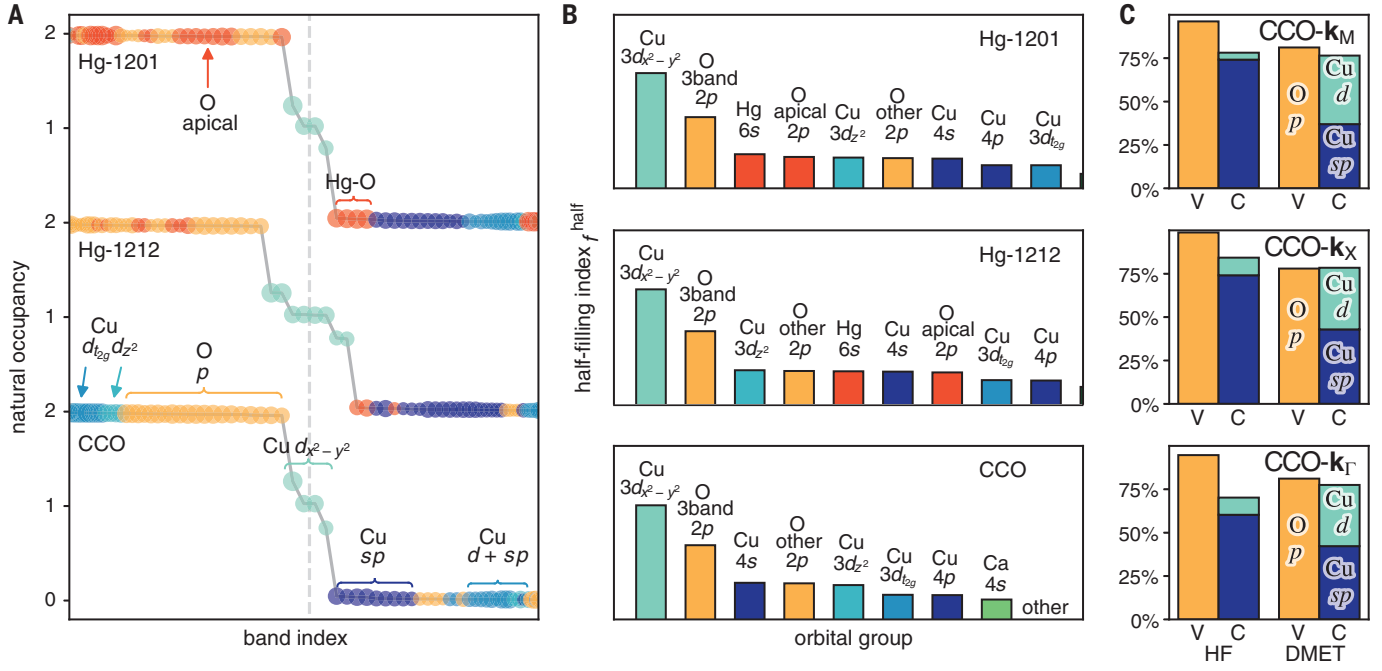


Fig. 3. Natural occupancy distribution (eigenvalues of the single-particle density matrix) and quasiparticle character. (**A**) Occupancy of natural orbitals around the Fermi level (dashed line), from the spin-traced density matrix ($\gamma^\alpha + \gamma^\beta$) in Hg-1201, Hg-1212, and CCO. Orbital character is indicated with colors and labels. (**B**) Half-filling index of the different local orbitals (definition is provided in eq. S50), measuring their importance in the most correlated orbitals of the calculation. (**C**) Orbital component analysis of the spin-resolved mean-field (HF) and correlated (DMET) top valence (V) and bottom conduction (C) bands of CCO at different **k** points (averaged from the eight bands near the Fermi level). Γ, (0, 0); X, $(\frac{1}{2}, 0)$; M, $(\frac{1}{2}, \frac{1}{2})$.

Natural occupancy distributions and effects of correlation

We obtain additional insight from the spin-resolved (γ^σ) and spin-traced ($\gamma = \sum_\sigma \gamma^\sigma$) single-particle density matrices (equal-time Green's functions) evaluated in the full crystal. These provide nonlocal and **k**-space information on correlations. We first discuss the spin-traced single-particle density matrix. The eigenvalues (the natural occupancy distribution, sometimes called the momentum distribution function) and eigenvectors (natural orbitals) illustrate the degree of symmetry breaking and highlight the important degrees of freedom near the Fermi level. The spin-traced natural occupancy distribution together with the projected atomic character of the eigenvectors is shown in Fig. 3, A and B. The most important orbitals near the Fermi level are the classic three-band orbitals: Cu $3d_{x^2-y^2}$ and O $2p_x$, $2p_y$. We also find no single next-most-important orbital; Cu $4s$, $4p$, $3d_{z^2}$, and the apical oxygen and Hg orbitals all contribute to a similar degree.

The spin-resolved natural occupancies and eigenvectors indirectly reflect the nature of the quasiparticles and the importance of dynamical effects [discussion is available in (*34*), section 1.3.4]. The eigenvectors with natural occupancies closest to the jump across the Fermi level can be viewed as "pseudo"–valence band maximum (VBM)–conduction band minimum (CBM) states. Defined in this way, the pseudo-VBM is dominated by O $2p_{x(y)}$, whereas the pseudo-CBM is dominated by Cu $3d_{x^2-y^2}$ (and the apical O and Hg bands in the Hg-Ba compounds) (Fig. 3C). This classifies all the compounds as charge-transfer insulators.

We further untangle the effect of interactions from pure single-particle physics by comparing the spin-resolved natural occupancies of the correlated calculation with that of a spin-polarized Hartree-Fock (HF) reference. The correlated spin-resolved natural occupancies are all quite close to 0 and 1 (fig. S15), the mean-field values; thus, dynamical effects are small. However, the orbital components of the eigenvectors are very different between the mean-field and correlated distributions (Fig. 3C), indicating strong static effects. It appears that in the AFM state the effect of interactions on the quasiparticles is mainly static rather than dynamical and can be captured largely through static screening of the interactions, correlation-driven rehybridization of the orbitals, and renormalization of their energies.

Magnetic trends across the cuprates

We next characterize the low-lying magnetic excitations across the series of cuprates. To do so compactly, we introduce a magnetic model (not to solve for the electronic structure, but for interpretation) and extract exchange couplings from our correlated calculations of different spin configurations: the AFM state, the FM state, and a spin-density wave state (fig. S3). From these, we derive parameters for the NN Heisenberg model (J) and a multi-J Heisenberg model in which the exchange couplings J_1, J_2, J_3, and J_c are related through the perturbation expansion of the one-band Hubbard model (with only three free parameters). [A $3J^{eff}$ model, in which J_c is renormalized into the J_1, J_2, J_3 parameters, can also be derived. In CCO, we also derive an interlayer $J_\perp$ using two AFM layer configurations. A full discussion of all models and the spin-wave calculation is available in (*34*) section 1.5.2.] The parameters are illustrated in Fig. 4B and tabulated in tables S10 to S13. We display the corresponding spin-wave spectrum from linear spin wave theory in Fig. 4A.

Spin-wave spectrum

In CCO, the full experimental dispersion is available, whereas for Hg-1201 and Hg-1212, only part of the dispersion near the Γ point has been measured. As expected, the NN Heisenberg model does not capture dispersion away from the Γ point; however, the derived NN J agrees well with that derived from experiment by fitting near the Γ point. For example, in CCO, the NN J fit to DMET data yields J = 155 meV, compared with J = 142,158 meV (the two numbers are from different experiments) (*47*, *48*). The multi-J model

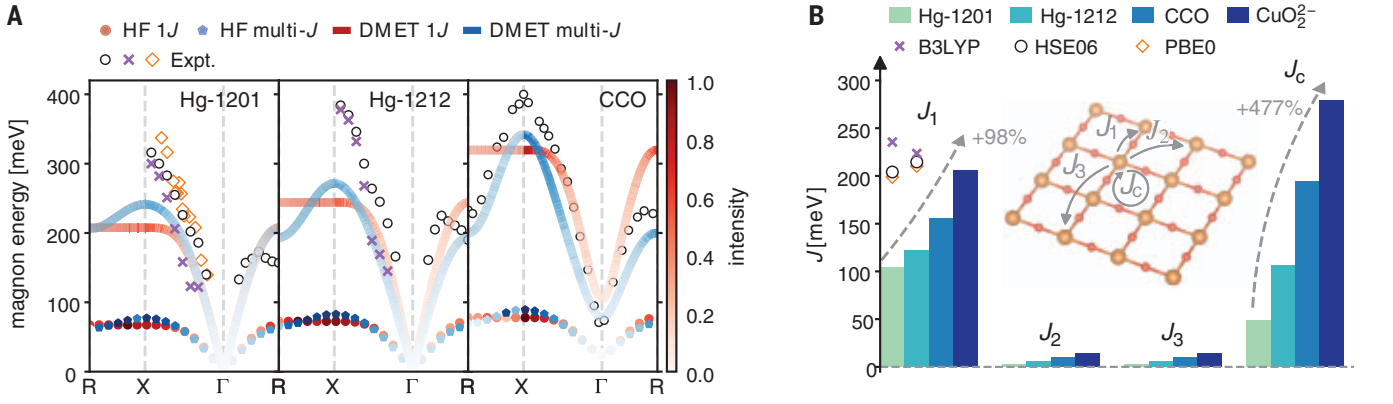


Fig. 4. Spin wave dispersion of Hg-1201, Hg-1212, and CCO. (**A**) The 2D magnetic Brillouin zone is sampled along Γ, (0, 0); X, $(\frac{1}{2}, 0)$; R, $(\frac{1}{4}, \frac{1}{4})$; k_z is fixed at 0.46 to match the experimental conditions in CCO and fixed at 0 for Hg-1201 and Hg-1212. NN Heisenberg (1 *J*) and multi-*J* model curves are shown. The multi-*J* model includes a quantum renormalization factor of 1.219 (*52*). Experimental RIXS data are extracted from (*49*, *53*) for Hg-1201 and Hg-1212 and (*47*) for CCO. (**B**) Trends in the multi-*J* model parameters across the cuprate family. Hybrid density functional (PBE0, HSE06, and B3LYP) results for the first two Hg compounds are also shown with symbols. Details are available in (*34*), section 2.3.

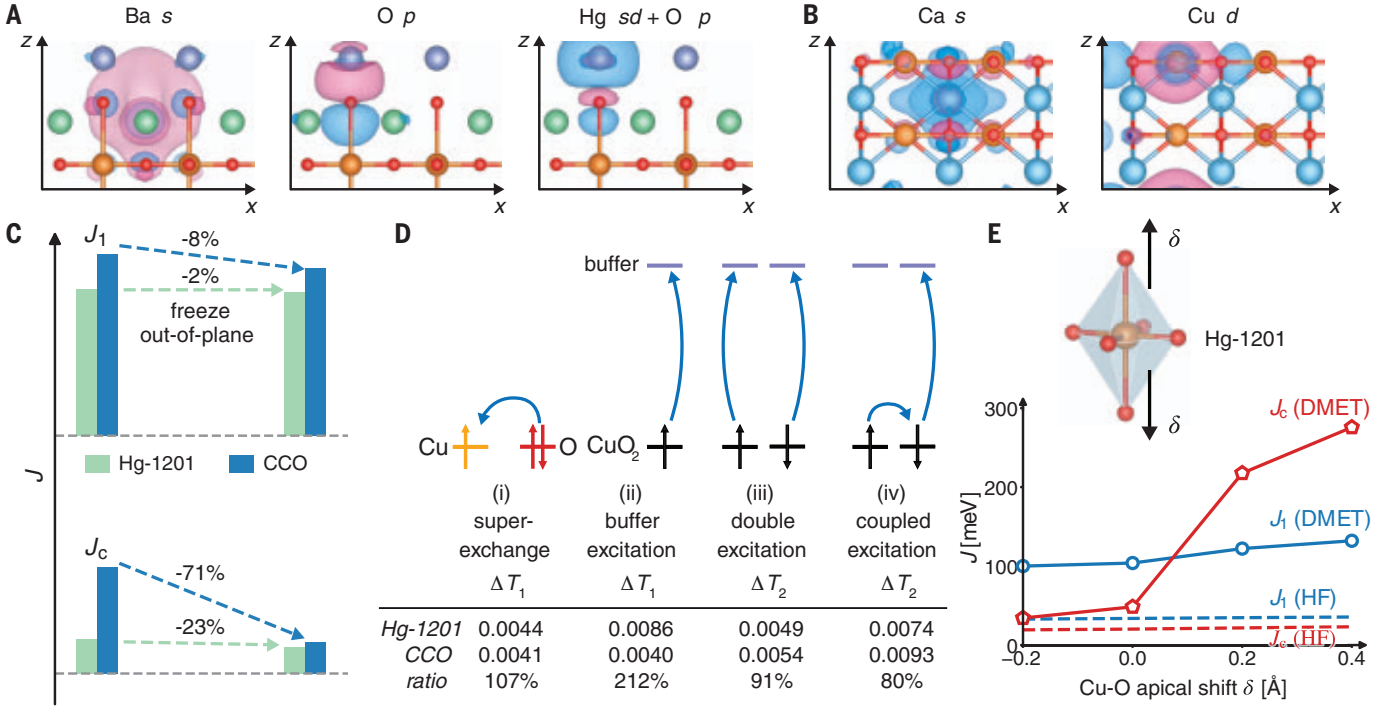


Fig. 5. Effects of buffer layers. (**A** and **B**) Representative out-of-plane orbitals (isosurfaces) in (A) Hg-1201 and (B) CCO. (**C**) The effect of freezing fluctuations to out-of-plane orbitals on the NN magnetic coupling J_1 and cyclic exchange coupling J_c. (**D**) Excitations relevant to exchange pathways in cuprates. (i) Superexchange is facilitated by excitations from in-plane oxygen orbitals to empty copper states; (ii) in Hg-1201, substantial excitations from the copper-oxygen plane to the buffer layer reduce superexchange. Double-excitation pathways (iii) and (iv) involving buffer layers are also shown. The numbers (ΔT_1, ΔT_2) reflect the change in excitation weight upon unfreezing the buffer orbitals. (**E**) Influence of apical Cu–O distance on exchange coupling J_1 and J_c at the mean-field (HF) and correlated (DMET) level.

with ab initio parameters yields improved agreement across the experimental dispersion, illustrating the importance of long-range exchange. The discrepancies are largest near the X point $(\frac{1}{4}, \frac{1}{4})$, likely owing to finite size effects in the embedding, although there are also confounding factors from the experimental setting in Hg-1201 and Hg-1212 [(*34*), section 2.3.5]. Compared with CCO, the Hg–Ba compounds display flatter dispersions, and we capture this in our derived spin-wave spectrum.

Magnetic parameters

Trends in the magnetic couplings of the multi-*J* Heisenberg model among the four compounds are shown in Fig. 4B. Across the series Hg-1201, Hg-1212, CCO, CuO_2^{2-}, all couplings J_1, J_2, J_3, and J_c increase substantially. J_1 roughly doubles, and J_c increases by a factor of 5, illustrating (i) the importance of the buffer layers in the long-range exchange coupling and (ii) the increasing "delocalization" across the series of compounds. A recent resonant inelastic x-ray scattering (RIXS) experiment (*49*) suggests that J_1 increases significantly (by about 20 to 30%) from the single-layer

Hg-1201 to the double-layer Hg-1212, which is similar to the increase in T_c. We find quantitative agreement with our correlated calculations, in which Hg-1212 shows an increase in J_1 by about 18%.

Effect of interactions

To understand the effect of interactions, we can compare to the mean-field HF results. These give almost flat dispersion curves because the J couplings are very small (for example, $J_1 \sim 40$ meV), and the magnon energy at the Γ point is also lower than the experimental value. Thus, the observed magnetic energy scales require a careful treatment of electron correlations. As suggested above, a large part of the effect of interactions can be captured by a renormalization of the low-energy band structure and interaction. Choosing a density functional treatment or Hubbard U parameter can mimic this; however, we do not find that a single choice of functional or Hubbard U consistently or accurately reproduces the material trends. For example, moving from Hg-1201 to Hg-1212 should yield a substantial increase in the exchange couplings, but J_1 decreases with the B3LYP functional and increases only marginally with HSE06 (5%) and PBE0 (6%) (Fig. 4B, symbol data). In addition, J_1 is overestimated by all of the above functionals.

Untangling layer effects

We now connect the microscopic correlated electronic structure with the trends in the magnetic physics observed above to derive mechanistic insights. Changing the buffer layer leads to large changes in the exchange couplings (particularly for the nonlocal terms). However, this effect does not appear at the HF mean-field level. To verify that it is caused by fluctuations (electron correlations) coupling the CuO_2 plane with the buffer layers (and not simply the effect of the electrostatic potential of the buffer layer on electron correlation within the cuprate plane), we devise a procedure that allows us to switch electron correlation with the buffer layer orbitals on and off. To do so, we explicitly freeze excitations involving out-of-plane orbitals in the correlated impurity solver calculations (the impurity wave function excludes particle-hole excitations from the HF determinant involving out-of-plane orbitals) [(*34*), section 2.4.1]. Any changes from freezing and unfreezing these fluctuations therefore directly reflect the influence of electron correlation with the orbitals of the buffer plane.

Representative out-of-plane orbitals of Hg-1201 and CCO are shown in Fig. 5, A and B. The out-of-plane impurity orbitals consist of empty outer valence shells on Ca, Hg, and Ba apical oxygen orbitals and other orbitals that originate from the adjacent copper-oxygen plane. The Ca- and Ba-centered localized orbitals (4*s* and 6*s*) are similar in CCO and Hg-1201.

The changes in the J_1 and J_c from unfreezing the out-of-layer orbitals are shown in Fig. 5C. In both compounds, the exchange couplings are decreased by freezing, but in CCO, the effect is stronger, and J_c is especially strongly influence by freezing, decreasing by as much as 71% in CCO. To understand this, we analyze the correlated impurity wave functions in CCO and Hg-1201. Shown in Fig. 5D are the changes in the weights of single-particle excitations ΔT_1 and connected two-particle excitations ΔT_2 upon unfreezing the buffer layer in the two compounds. Generally speaking, when the buffer layer is unfrozen, the increased excitation manifold increases screening and decreases the energetic penalty to excite from filled to empty states, such as the empty Cu and buffer-layer states. In CCO and Hg-1201, we find that this increases the O → Cu excitation associated with super-exchange [Fig. 5D, process (i)], increasing the exchange couplings. However, in Hg-1201, we see in addition a significant increase in excitations from in-plane Cu, O orbitals to the empty Hg, apical O states [Fig. 5D, process (ii)]. This change in the copper plane to buffer excitation is more than twice as large in Hg-1201 as in CCO, and it depletes the ground configuration associated with in-plane exchange and reduces the effective nonlocal hopping by rehybridizing the Cu empty states (*16*), canceling the enhancement of in-plane O → Cu excitations and yielding an aggregate small change in exchange coupling upon unfreezing the buffer orbitals [(*34*), section 2.4.2]. This also explains why the exchange couplings of Hg-1212 lie in between those of Hg-1201 and CCO: The buffer suppression of in-plane super-exchange occurs through a single buffer layer in Hg-1212 versus two on either side in Hg-1201. The analysis also reveals (smaller) differences between the compounds in the connected two-particle fluctuations involving the buffer [Fig. 5D, processes (iii) and (iv)]; these are material-specific effects that cannot be folded into a static renormalization. Last, in Fig. 5E we show the effect on the exchange coupling of increasing the apical oxygen distance in Hg-1201, both at the mean-field level and at the correlated level. Consistent with the above mechanism, we find that increasing apical oxygen distance removes the buffer suppression effect in the correlated calculation (increasing the exchange coupling) but makes little difference in the mean-field calculation because fluctuations must first renormalize the energies of the empty states for them to be accessible.

Discussion

We have demonstrated that through a numerical strategy that combines quantum embedding, ab initio quantum solvers, and periodic quantum chemistry, we can determine material-specific correlated electron structure in the parent state of the cuprates at the many-body level. This reveals trends in the multi-orbital bonding and correlation effects in the Fermi distribution and quasiparticles and gives a quantitative description of the low-energy magnetic excitations. Across a series of homologous mercury-barium and calcium cuprates, the systematic trends in the nature of the magnetic exchange can be explained through the analysis of the many-body state, which uncovers a competition between superexchange and plane-to-buffer excitation processes.

A general observation is that although the interactions are strong, many of their effects in the parent state can be renormalized into a static low-energy theory. This supports the long-standing practice of interpreting physics in this region through simple band structures and static interaction parameters. However, we also find that empirical approaches to determine this renormalization do not have the accuracy to capture the trends among the materials, unlike the controlled many-body approaches used here.

A strength of the many-body approach is that we can interrogate individual electronic processes, and our ab initio formulation allows us to trace these processes beyond models to the individual atomic orbital level. We used this capability to untangle the links between layer composition and magnetic exchange. In prescient work, it was conjectured that the range of magnetic exchange is related to electronic processes involving an effective apical conduction band and that this further correlates with the superconducting transition temperature (*16*). We now have a direct picture of the first part of this conjecture, with rich atomic-scale and many-body resolution.

Components of the numerical strategy in this work have previously been used to describe exotic phases in models. The success of the current ab initio realization for cuprate parent states thus gives hope that a similar approach may eventually yield a quantitative picture of more complex cuprate phases. If that is the case, we may be able to answer the second part of the above and similar conjectures about superconducting properties, through a direct ab initio simulation of the superconducting orders and the energy scales of the cuprates in their doped states.

REFERENCES AND NOTES

1. E. Dagotto, *Science* **309**, 257–262 (2005).
2. R. M. Martin, L. Reining, D. M. Ceperley, *Interacting Electrons* (Cambridge Univ. Press, 2016).
3. E. Dagotto, *Rev. Mod. Phys.* **66**, 763–840 (1994).
4. J. Orenstein, A. J. Millis, *Science* **288**, 468–475 (2000).
5. N. Plakida, *High-Temperature Cuprate Superconductors: Experiment, Theory, and Applications*, vol. 166 (Springer, 2010).
6. X. Zhou *et al.*, *Nat. Rev. Phys.* **3**, 462–465 (2021).
7. A. Schilling, M. Cantoni, J. D. Guo, H. R. Ott, *Nature* **363**, 56–58 (1993).
8. H. Zhai, G. K.-L. Chan, *J. Chem. Phys.* **154**, 224116 (2021).

9. T. Zhu, C. A. Jiménez-Hoyos, J. McClain, T. C. Berkelbach, G. K.-L. Chan, *Phys. Rev. B* **100**, 115154 (2019).
10. Z.-H. Cui, T. Zhu, G. K.-L. Chan, *J. Chem. Theory Comput.* **16**, 119–129 (2020).
11. T. Zhu, Z.-H. Cui, G. K.-L. Chan, *J. Chem. Theory Comput.* **16**, 141–153 (2020).
12. T. Zhu, G. K.-L. Chan, *Phys. Rev. X* **11**, 021006 (2021).
13. Q. Sun, T. C. Berkelbach, J. D. McClain, G. K.-L. Chan, *J. Chem. Phys.* **147**, 164119 (2017).
14. Q. Sun *et al.*, *J. Chem. Phys.* **153**, 024109 (2020).
15. B.-X. Zheng *et al.*, *Science* **358**, 1155–1160 (2017).
16. E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, O. K. Andersen, *Phys. Rev. Lett.* **87**, 047003 (2001).
17. E. Antipov, A. Abakumov, S. Putilin, *Supercond. Sci. Technol.* **15**, R31–R49 (2002).
18. K. Lokshin *et al.*, *Phys. Rev. B Condens. Matter* **63**, 064511 (2001).
19. H. Sakakibara, H. Usui, K. Kuroki, R. Arita, H. Aoki, *Phys. Rev. Lett.* **105**, 057003 (2010).
20. C. Weber, C. Yee, K. Haule, G. Kotliar, *Europhys. Lett.* **100**, 37001 (2012).
21. S. Kim, X. Chen, W. Fitzhugh, X. Li, *Phys. Rev. Lett.* **121**, 157001 (2018).
22. L. K. Wagner, P. Abbamonte, *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 125129 (2014).
23. K. Foyevtsova *et al.*, *Phys. Rev. X* **4**, 031003 (2014).
24. M. S. Hybertsen, E. B. Stechel, M. Schluter, D. R. Jennison, *Phys. Rev. B Condens. Matter* **41**, 11068–11072 (1990).
25. D. Muñoz, F. Illas, I. de P. R. Moreira, *Phys. Rev. Lett.* **84**, 1579 (2000).
26. C. Lane *et al.*, *Phys. Rev. B* **98**, 125140 (2018).
27. P. R. C. Kent *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **78**, 035132 (2008).
28. C. Weber, K. Haule, G. Kotliar, *Nat. Phys.* **6**, 574–578 (2010).
29. S. R. White, D. J. Scalapino, *Phys. Rev. B Condens. Matter Mater. Phys.* **92**, 205112 (2015).
30. S. Choi, A. Kutepov, K. Haule, M. Van Schilfgaarde, G. Kotliar, *npj Quant. Mater.* **1**, 16001 (2016).
31. M. Qin *et al.*, *Phys. Rev. X* **10**, 031016 (2020).
32. J. Karp *et al.*, *Phys. Rev. X* **10**, 021061 (2020).
33. P. Mai, G. Balduzzi, S. Johnston, T. A. Maier, *npj Quant. Mater.* **6**, 26 (2021).
34. Methods are available in the supplementary materials.
35. Q. Sun, G. K.-L. Chan, *Acc. Chem. Res.* **49**, 2705–2712 (2016).
36. A. Georges, G. Kotliar, W. Krauth, M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13–125 (1996).
37. G. Kotliar *et al.*, *Rev. Mod. Phys.* **78**, 865–951 (2006).
38. A. A. Kananenka, E. Gull, D. Zgid, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 121111 (2015).
39. G. Knizia, G. K.-L. Chan, *Phys. Rev. Lett.* **109**, 186404 (2012).
40. S. Wouters, C. A. Jiménez-Hoyos, Q. Sun, G. K.-L. Chan, *J. Chem. Theory Comput.* **12**, 2706–2719 (2016).
41. B.-X. Zheng, J. S. Kretchmer, H. Shi, S. Zhang, G. K.-L. Chan, *Phys. Rev. B* **95**, 045103 (2017).
42. P. Werner, A. J. Millis, *Phys. Rev. Lett.* **104**, 146401 (2010).
43. J. R. Bartlett, M. Musiał, *Rev. Mod. Phys.* **79**, 291–352 (2007).
44. A. Shee, D. Zgid, *J. Chem. Theory Comput.* **15**, 6010–6024 (2019).
45. S. R. White, R. L. Martin, *J. Chem. Phys.* **110**, 4127–4130 (1999).
46. F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
47. Y. Peng *et al.*, *Nat. Phys.* **13**, 1201–1206 (2017).
48. D. Kan, A. Yamanaka, T. Terashima, M. Takano, *Physica C* **412–414**, 298–302 (2004).
49. L. Wang *et al.*, *Nat. Commun.* **13**, 1 (2022).
50. Z.-H. Cui, zhcui/cuprate_parent_state_data: arXiv.2112.09735. Zenodo (2022).
51. Z.-H. Cui, gkclab/libdmet_preview: arXiv:2112.09735. Zenodo (2022).
52. J.-Y. Delannoy, M. Gingras, P. Holdsworth, A.-M. Tremblay, *Phys. Rev. B Condens. Matter Mater. Phys.* **79**, 235130 (2009).
53. B. Yu *et al.*, *Phys. Rev. X* **10**, 021059 (2020).

ACKNOWLEDGMENTS

We thank T. Zhu, L. Peng, Y. Li, P. Lee, A. Millis, and S. White for helpful discussions. **Funding:** This work was primarily supported by the US Department of Energy, Office of Science, through grant DE-SC0018140. The DMRG calculations were performed using the BLOCK2 code, which was developed with funding from the US National Science Foundation through grant CHE-2102505. This work also relied on improvements to the PYSCF density fitting and integral transformation modules, carried out as part of work supported by the Center for Molecular Magnetic Quantum Materials, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences, under award DE-SC0019330. G.K.-L.C. is a Simons Investigator in Physics and is part of the Simons Collaboration on the Many-Electron Problem. Z.-H.C. acknowledges support from the Eddleman Quantum Institute through a graduate fellowship. Calculations were conducted in the Resnick High Performance Computing Center, supported by the Resnick Sustainability Institute at Caltech, and National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory. **Author contributions:** Z.-H. C. and G.K.-L. C. designed the study and wrote the manuscript. Z.-H.C. and H.Z. performed the calculations. X.Z. contributed to the **k** point symmetry code of PYSCF and LIBDMET. Z.-H.C., H.Z., and G.K.-L.C. discussed the results. All authors contributed to the writing and editing of the manuscript. **Competing interests:** G.K.-L.C. is a part owner of Qsimulate. **Data and materials availability:** Data used in this work are in the supplementary materials and online at https://github.com/zhcui/cuprate_parent_state_data and (*50*). The LIBDMET code is available at https://github.com/gkclab/libdmet_preview and (*51*). The BLOCK2 code is available at https://github.com/block-hczhai/block2-preview. PYSCF is available from https://pyscf.org.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm2295
Supplementary Text
Figs. S1 to S24
Tables S1 to S18
References (*54–111*)

Submitted 2 September 2021; accepted 28 July 2022
10.1126/science.abm2295

REPORTS

MESOSCOPIC PHYSICS

Isolated ballistic non-abelian interface channel

Bivas Dutta[1], Vladimir Umansky[1], Mitali Banerjee[2], Moty Heiblum[1]*

The topological order of a quantum Hall state is mirrored by the gapless edge modes owing to bulk-edge correspondence. The state at the filling of ν = 5/2, predicted to host non-abelian anyons, supports a variety of edge modes (integer, fractional, neutral). To ensure thermal equilibration between the edge modes and thus accurately determine the state's nature, it is advantageous to isolate the fractional channel (1/2 and neutral modes). In this study, we gapped out the integer modes by interfacing the ν = 5/2 state with integer states ν = 2 and ν = 3 and measured the thermal conductance of the isolated-interface channel. Our measured half-quantized thermal conductance confirms the non-abelian nature of the ν = 5/2 state and its particle-hole Pfaffian topological order. Such an isolated channel may be more amenable to braiding experiments.

The fractional quantum Hall effect harbors fractionally charged quasiparticles localized in the two-dimensional (2D) bulk, surrounded by gapless chiral edge modes at the periphery (*1*). Laughlin's states and their particle-hole conjugated states are abelian (*1*). In higher Landau levels, exotic non-abelian states are expected. When the bulk hosts a large number of well-separated quasiparticles, the ground state is highly degenerate (*2*, *3*). A proposed state that hosts such exotic quasiparticles is the ν = 5/2 state, which supports gapless edge modes: two integer modes (ν = 2), a fractional mode (ν = 1/2), and neutral Majorana modes (their number and chirality depend on the state's topological order). To capture the nature of the state, we recently employed thermal conductance (G_{th}) measurements (*4*) of the edge modes, which are sensitive to all energy-carrying modes (charged and neutral alike). If all edge modes are fully thermally equilibrated, the measured G_{th} corresponds to a vital topological invariant that dictates the state's topological order.

We start by recapping the rationale for measuring the thermal conductance. In the abelian regime, with fully equilibrated downstream (DS) and upstream (US) modes, the two-terminal thermal conductance is given by $G_{th} = |n_d - n_u|\kappa_0 T$, where $\kappa_0 = \pi^2 k_B^2/3h$ is the quantum of thermal conductance, k_B is the Boltzmann constant, h is Planck's constant, T is temperature, and n_d (n_u) is the number of downstream (upstream) edge modes. However, in a complete absence of thermal equilibration, one expects $G_{th} = (n_d + n_u)\kappa_0 T$ (*5*). In the non-abelian regime, G_{th} is a fractional multiple of $\kappa_0 T$, originating from the fractional nature of the chiral central charge (*6*). With the assumption of full thermal equilibration among the multiple modes of the ν = 5/2 state, we experimentally found $K_{5/2} \cong 2.5\kappa_0$ (*4*), suggesting the particle-hole Pfaffian (PH-Pf) topological order. This result disagrees with numerical calculations (*7–22*).

In this work, we used a technique introduced in (*23*) to interface the ν = 5/2 state with the ν = 2 or ν = 3 states. The resulting ν = 1/2 isolated interface channel is composed of neutral mode(s) tied with the ν = 1/2 charge mode. The simpler structure of the isolated channel (without the

[1]Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Institute of Physics, Faculté of Basic Sciences, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.
*Corresponding author. Email: moty.heiblum@weizmann.ac.il

integer modes) eases considerably the intermode thermal equilibration (*4*). Our measured upstream noise of the interface channel, done in the thermally unequilibrated regime (*23*), made it possible to distinguish between the suggested topological orders, however, spontaneous edge reconstruction might add ambiguity. Here, measuring the thermal conductance coefficient of the interface channel in the fully thermally equilibrated regime, and finding $K_{1/2} = 0.5\kappa_0$, we prove the bulk's non-abelian nature and reaffirm its PH-Pf topological order (*4*, *23*). The isolated 1/2 channel, with its chiral neutral Majorana mode, is an excellent candidate for future interference experiments (i.e., braiding).

Molecular beam epitaxy (MBE)–grown high-quality GaAs-AlGaAs heterostructures, with shallow-DX-center doping (*24*), allow hysteresis-free gating and feature negligible heat conductance of the bulk (at millikelvin temperatures) (*23*). This doping method leads to higher disorder than in the conventional short-period superlattice doping technique (*4*, *24*, *25*). Our experimental setup containing a "two-arm" structure is shown in Fig. 1A. The inner gates (beige) and the outer ones (violet) tune the filling factor of each of the corresponding bulks in the sample. The gates are separated from the sample by a 25-nm-thick HfO_2 layer [for more details, see section II of the supplementary text (*26*)]. Gate voltage in the range $-1.5\ V < V_g < +0.3\ V$ allows varying the electron density from pinch-off to $3 \times 10^{11}\ cm^{-2}$ (fig. S1), controlling the filling factor over a wide range. The interface modes between the two bulks split when leaving the floating ohmic contact (20 μm by 2 μm, shown in red), into the two arms. This contact serves as a heated reservoir. Large ohmic contacts [source (S), drain (D), and ground (G), shown in cyan] are placed at the interfaces; separate contacts, located at the physical edge of the mesa (not shown), probe the filling factor of each of the respective bulks.

Current I_S injected from S toward the floating contact, carrying power $P_S = \frac{1}{2} I_S^2 R_S$, dissipates a power $\Delta P = \frac{1}{4} I_S^2 R_S$ in the contact (R_S is the resistance of the interface channel). Figure 1B provides a schematic representation of the heat balance in the floating contact at the interface. The heat leaving the contact is composed of a phononic contribution $\Delta P_{ph} = \beta\left(T_m^5 - T_0^5\right)$ and an electronic contribution $\Delta P_e = 0.5K\left(T_m^2 - T_0^2\right)$, where T_m is the temperature of the floating contact and T_0 is the base temperature, with $\Delta P = \Delta P_{ph} + \Delta P_e$ at equilibrium. For our small reservoir, the phononic contribution is negligible at temperatures $T_m < 20$ mK. The temperature T_m is determined by measuring the low-frequency Johnson-Nyquist (J-N) noise at the drain contact D, which is placed ~160 μm away along the downstream direction, in one of the two arms [see section III of the supplementary text (*26*)]. The noise is filtered by a resonant circuit (resonance frequency: 630 kHz; bandwidth: 10 to 30 kHz) and amplified by a low-noise cold voltage preamplifier (placed on the 4.2 K plate) followed by a room-temperature amplifier.

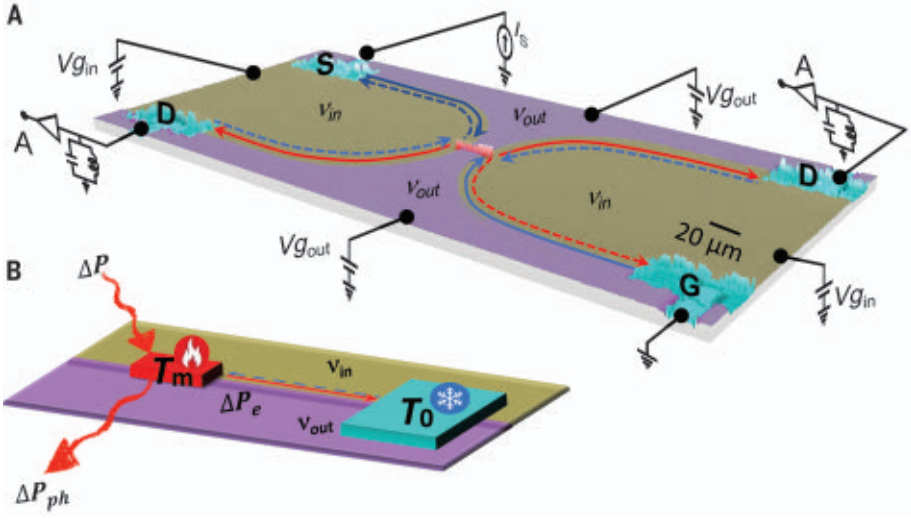


Fig. 1. The experimental setup used to measure the interface thermal conductance. (**A**) False colored scanning electron micrograph of the heart of the device. Two biased top gates, the inner gate (Vg_{in}, beige) and the outer gate (Vg_{out}, violet), divide the 2D electron gas (2DEG) mesa into two gate-defined arms. A small floating ohmic contact (red; dimensions: 20 μm by 2 μm) connects the two arms. Large contacts (S, D, and G, in cyan) probe the interface filling. Added ohmic contacts at the edge of the mesa (not shown) probe the filling of the respective sides. The floating contact is heated to T_m by an injecting current I_S from S. Its temperature is determined by measuring the low-frequency Johnson-Nyquist noise at 630 kHz (LC bandwidth: 10 to 30 KHz) after amplification by a cooled preamplifier (at 4.2 K) followed by a room-temperature amplifier. Arrows indicate the interface modes at the 5/2-3 interface as an example (see Fig. 4). (**B**) Schematic representation of heat balance in the floating contact, with the same color coding as in (A), showing only a zoomed-in part of the interface along the downstream direction. An injected current dissipates power ΔP, which is evacuated by two main channels: the phononic channel ΔP_{ph} and the electronic channel ΔP_e. The cold reservoir contact G is at temperature T_0. Because of the small volume of the floating contact, the phononic contribution is negligible at low temperatures.

In general, our experimental strategy is to tune the inner regions to the "tested" state ν_{in} (e.g., 5/2) and the outer regions to an integer state ν_{out} (e.g., 0, 1, 2, 3), leading to an interface filling $\nu_{int} = \nu_{in} - \nu_{out}$ (the nomenclature "$\nu_{in} - \nu_{out}$" is used in all figures). Note that the chirality of the interface charge mode reverses when the interfacing condition changes from $\nu_{in} > \nu_{out}$ to $\nu_{in} < \nu_{out}$. Owing to a single amplifier located downstream, the magnetic field was reversed between the latter two cases [see section III of the supplementary text (*26*)].

We start with measurements of the interface modes formed in the configurations 3-2, 3-1, and 3-0. The two-terminal S-D resistance, R_S, exhibits quantized plateaus as a function of the gate voltage, h/e^2, $h/2e^2$, and $h/3e^2$, with the expected thermal conductances $1\kappa_0 T$, $2\kappa_0 T$, and $3\kappa_0 T$, respectively (*27*, *28*). In these chiral configurations, the downstream modes are thermally equilibrated right from the start at the reservoir temperature T_m. The latter is determined by the measured J-N noise, S_{th}, at the drain [Fig. 2, A and B, and section V of the supplementary text (*26*)]. We analyzed the data in two ways: (i) by fitting the linear electronic contribution, ΔP versus T_m^2 for $T_m < 18$ mK, finding the thermal conductance of the interface $K \cong 1\kappa_0$, $2\kappa_0$, and $3\kappa_0$, for the three configurations, respectively (Fig. 2C); and (ii) by fitting the data for $T_m < 30$ mK (Fig. 2D), which reflects the electronic and phononic contributions. Using $\beta \approx 5 \times 10^{-9}\ W\ K^{-5}$, a similar quantization of the thermal conductance is found (*4*).

Interfacing the fillings $\nu = 7/3$, 5/2, and 8/3 with $\nu = 2$ and $\nu = 3$ leads to the source resistance shown in Fig. 3A. Well-quantized resistance plateaus indicate full charge equilibration at the interface. We start with the $\nu = 7/3$ and $\nu = 8/3$ abelian states. Interfacing these two states with $\nu = 2$ results in an effective mode filling of $\nu = 1/3$ for 7/3-2 and $\nu = 2/3$ for 8/3-2 (the latter is accompanied by an upstream neutral mode). The chirality at the interface is reversed when these states are interfaced with $\nu = 3$; with $\nu = 2/3$ charge channel for 7/3-3 and $\nu = 1/3$ mode for 8/3-3. The measured two-terminal interface thermal conductance for each of these four cases is shown in Fig. 3, B and C. The results agree with good accuracy with the theoretical expectations (*4*).

The $\nu = 5/2$ state may host different topological orders, with each having a different

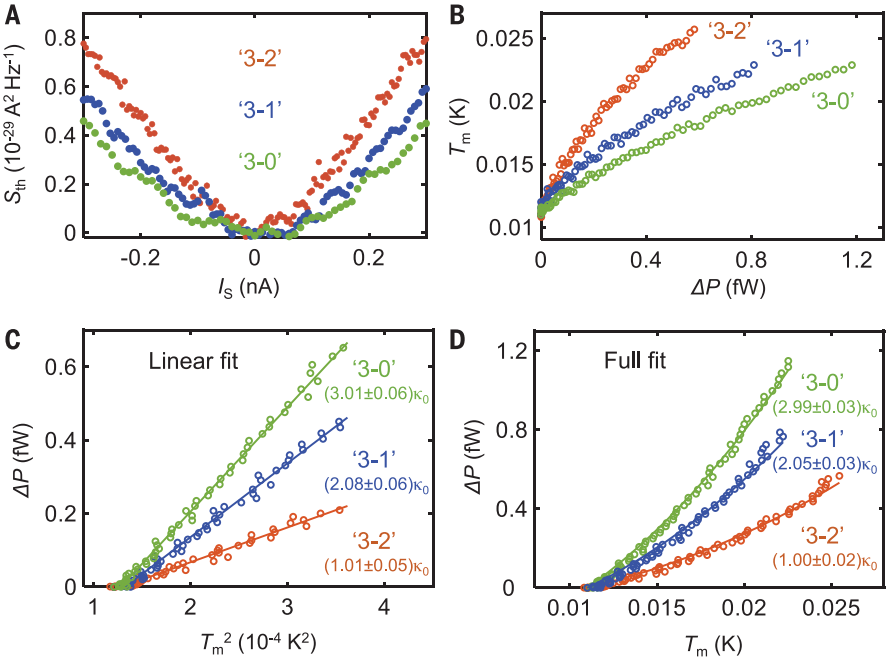


Fig. 2. Noise, temperature, and dissipation at integer interfaces. (**A**) Excess Johnson-Nyquist noise S_{th} as a function of heating current I_S for the three integer interfaces, 3-2 (orange), 3-1 (blue), and 3-0 (green), at T_0 = 11 mK. (**B**) Calculated temperature T_m as a function of the dissipated power $\Delta P = 0.25 I_S^2 R_S$, where R_S is the interface-mode resistance, for the three different configurations [see section V of the supplementary text (*26*)]. (**C**) Dissipated power as a function of squared temperature in the range $T_m < 18$ mK. The thermal conductance K is determined from the slope, via $\Delta P = \frac{1}{2}K(T_m^2 - T_0^2)$ and indicated in the figure. (**D**) Dissipated power as a function of temperature in the range $T_m < 30$ mK. Here, phonon contribution is included via $\Delta P = \frac{1}{2}K(T_m^2 - T_0^2) + \beta(T_m^5 - T_0^5)$, with electron–phonon coupling constant $\beta \approx 5 \times 10^9$ W K^{-5}; the extracted values of K are indicated in the figure.

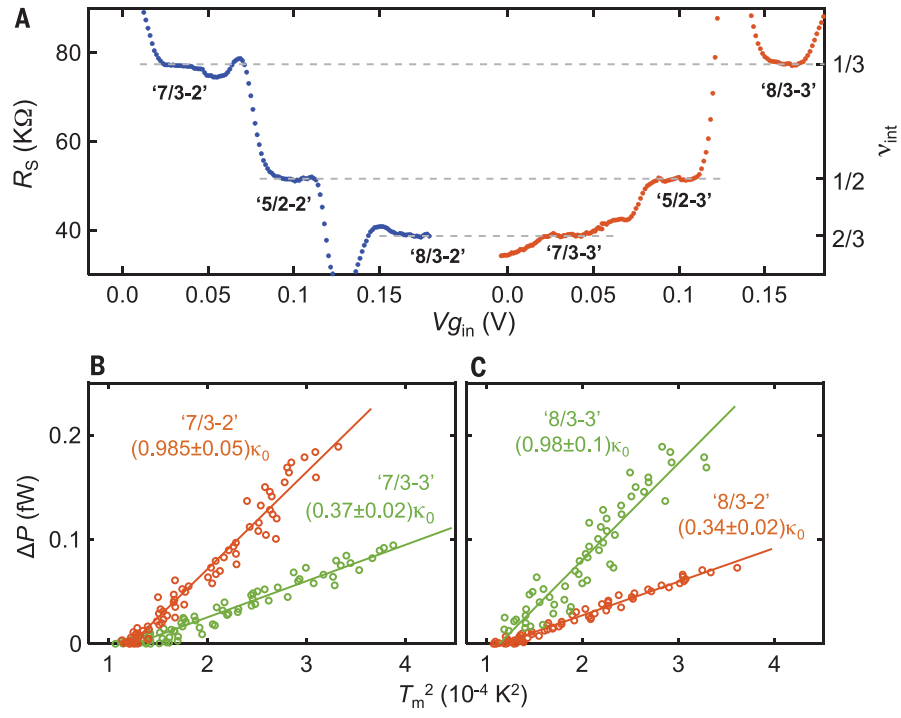


Fig. 3. Interface resistance and thermal conductance of states in the second Landau level. (**A**) Two terminal interface-mode resistance of 7/3-2 and -3, 5/2-2 and -3, and 8/3-2 and -3. Accurate quantized plateaus indicate full charge equilibration at the interface. The peaks and dips are attributed to reentrant effects (*30*). (**B**) Determination of the thermal conductance for $T_m < 18$ mK, as was done in Fig. 2. The slope for 7/3-2 is close to unity, as expected for the 1/3 interface mode. The slope for 7/3-3 is 0.37κ_0, not zero, owing to the well-known lack of equilibration of the resultant 2/3 mode (*27*). (**C**) Similar data as in (B), at different interfacing conditions with ν = 8/3.

mode structure and a different thermal conductance (*29*). Here, we consider two competing candidates: particle-hole Pfaffian (PH-Pf) and anti-Pfaffian (A-Pf). For a discussion of other possible orders, see section VI of the supplementary text (*26*). Both orders support counterpropagating modes at the bare edge of the sample, namely, the 5/2-0 configuration (Fig. 4, A and B). Here, ν = 0 reflects the outside vacuum. Aside from two downstream integer charge modes and a ν = 1/2 charge mode, the PH-Pf order supports an upstream Majorana mode, whereas the A-Pf order supports three upstream Majorana modes (two Majorana modes form a single bosonic neutral mode). For PH-Pf, this leads to a thermal conductance of $2.5\kappa_0 T$ for all thermally equilibrated modes and $3.5\kappa_0 T$ for thermally unequilibrated modes; for A-Pf, the values are $1.5\kappa_0 T$ for thermally equilibrated and $4.5\kappa_0 T$ for thermally unequilibrated. Thus, with counterpropagating modes, an accurate determination of the topological order on the basis of thermal conductance requires a full thermal equilibration between all modes.

Figure 4, A to F, presents different configurations: 5/2-0, 5/2-1, and 5/2-2. As shown in the figures, the expected thermal conductance of the two competing orders may overlap when considering the range from a full thermal equilibration to no equilibration. However, interfacing 5/2-3 clearly distinguishes between the topological orders (Fig. 4, G and H). In an A-Pf

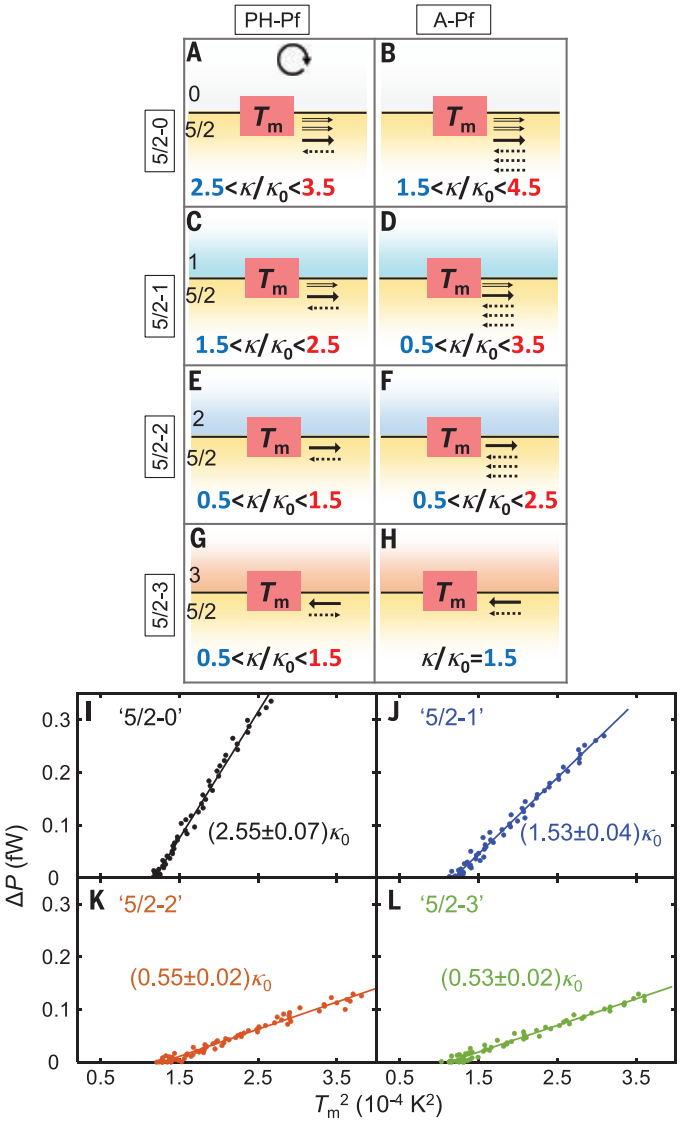


Fig. 4. Thermal conductance of the 5/2-*n* interfaces. (**A** to **H**) Plots of the interface modes of the interfaced PH-Pf and the A-Pf orders with different integers and the expected thermal conductance in each case. Fully equilibrated value in light blue and unequilibrated value in red. The double-line arrow indicates an integer mode, the single-line arrow indicates the 1/2 charge mode, and the dashed-line arrow indicates the Majorana mode. Clockwise chirality is indicated by the circled arrow. (**I** to **L**) Plots of the dissipated power as a function of the squared temperature with linear fits as in Figs. 2 and 3, for all four 5/2 interfaces. The extracted thermal conductance values, in particular for 5/2-3, unambiguously exclude the A-Pf topological order [see (G) and (H)] and agrees (with small estimated errors) with the Ph-Pf topological order.

order, the interface channel supports two co-propagating modes, which leave the hot reservoir at the same temperature, T_m (hence, equilibrated), leading to a thermal conductance of $1.5\kappa_0 T$. By contrast, the PH-Pf order supports an interface channel with counter-propagating charge and Majorana mode, leading to thermal conductance in the range $0.5\kappa_0 T$ to $1.5\kappa_0 T$.

In Fig. 4, I to L, we present the measurement results of ΔP versus T_m^2 for the four 5/2 interfaces: 5/2-0, 5/2-1, 5/2-2, and 5/2-3. The measured thermal conductances are $(2.55 \pm 0.07)\kappa_0 T$, $(1.53 \pm 0.04)\kappa_0 T$, $(0.55 \pm 0.02)\kappa_0 T$, and $(0.53 \pm 0.02)\kappa_0 T$, respectively. These results, especially for the decisive 5/2-3 interface, rule out the A-Pf order and support the PH-Pf order (Fig. 4, A to H). Notably, the observed 1/2-quanta thermal conductance of the isolated 1/2-channel at the 5/2-3 interface establishes the non-abelian nature of the bulk.

The $\nu = 5/2$ state has been attracting widespread attention owing to growing expectations of realizing non-abelian anyons in condensed matter systems. Here, we exploit the interfacing method introduced in (*23*), which allows measuring the thermal conductance of an emergent interface channel at the interface between two adjacent quantum Hall states ($\nu = 5/2$ and integers $\nu = 2$ and $\nu = 3$). The advantage of the interfacing method lies in the simplification of equilibration among the remaining modes, providing a direct determination of the non-abelian nature of the bulk, along with a clear distinction between competing orders of the $\nu = 5/2$ state. The isolated 1/2-channel is amenable to interference experiments, which could test its robustness against decoherence mechanisms [see section IX of the supplementary text (*26*)].

REFERENCES AND NOTES

1. M. Heiblum, D. E. Feldman, *Int. J. Mod. Phys. A* **35**, 2030009 (2020).
2. X. G. Wen, *Phys. Rev. Lett.* **66**, 802–805 (1991).
3. J. B. Pendry, *J. Phys. Math. Gen.* **16**, 2161–2171 (1983).
4. M. Banerjee *et al.*, *Nature* **559**, 205–210 (2018).
5. C. L. Kane, M. P. A. Fisher, *Phys. Rev. B Condens. Matter* **55**, 15832–15837 (1997).
6. A. Gromov, G. Y. Cho, Y. You, A. G. Abanov, E. Fradkin, *Phys. Rev. Lett.* **114**, 016805 (2015).
7. R. H. Morf, *Phys. Rev. Lett.* **80**, 1505–1508 (1998).
8. M. Storni, R. H. Morf, S. Das Sarma, *Phys. Rev. Lett.* **104**, 076803 (2010).
9. G. Moore, N. Read, *Nucl. Phys. B* **360**, 362–396 (1991).
10. M. Levin, B. I. Halperin, B. Rosenow, *Phys. Rev. Lett.* **99**, 236806 (2007).
11. S. S. Lee, S. Ryu, C. Nayak, M. P. Fisher, *Phys. Rev. Lett.* **99**, 236807 (2007).
12. M. P. Zaletel, R. S. K. Mong, F. Pollmann, E. H. Rezayi, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 045115 (2015).
13. E. H. Rezayi, *Phys. Rev. Lett.* **119**, 026801 (2017).
14. P. T. Zucker, D. E. Feldman, *Phys. Rev. Lett.* **117**, 096802 (2016).
15. D. F. Mross, Y. Oreg, A. Stern, G. Margalit, M. Heiblum, *Phys. Rev. Lett.* **121**, 026801 (2018).
16. C. Wang, A. Vishwanath, B. I. Halperin, *Phys. Rev. B* **98**, 045112 (2018).
17. I. C. Fulga, Y. Oreg, A. D. Mirlin, A. Stern, D. F. Mross, *Phys. Rev. Lett.* **125**, 236802 (2020).
18. B. Lian, J. Wang, *Phys. Rev. B* **97**, 165124 (2018).
19. S. H. Simon, *Phys. Rev. B* **97**, 121406 (2018).
20. D. E. Feldman, *Phys. Rev. B* **98**, 167401 (2018).
21. S. H. Simon, B. Rosenow, *Phys. Rev. Lett.* **124**, 126801 (2020).
22. K. K. W. Ma, D. E. Feldman, *Phys. Rev. Lett.* **125**, 016801 (2020).
23. B. Dutta *et al.*, *Science* **375**, 193–197 (2022).
24. V. Umansky, M. Heiblum, in *Molecular Beam Epitaxy: From Research to Mass Production*, M. Henini, Ed. (Elsevier Science BV, 2013), pp. 121–137.
25. K. A. Villegas Rosales *et al.*, *Phys. Rev. Lett.* **127**, 056801 (2021).
26. See supplementary materials.
27. M. Banerjee *et al.*, *Nature* **545**, 75–79 (2017).
28. S. Jezouin *et al.*, *Science* **342**, 601–604 (2013).
29. K. K. W. Ma, D. E. Feldman, *Phys. Rev. B* **100**, 035302 (2019).
30. M. O. Goerbig, P. Lederer, C. Morais Smith, *Phys. Rev. B Condens. Matter* **68**, 241302 (2003).
31. B. Dutta, V. Umansky, M. Banerjee, M. Heiblum, Isolated ballistic non-abelian interface channel, Zenodo (2022); https://doi.org/10.5281/zenodo.6559780.

ACKNOWLEDGMENTS

We acknowledge D. Mross for a critical reading of the manuscript and A. Stern for valuable suggestions. **Funding:** B.D. acknowledges support from the Clore Foundation. M.H. acknowledges the continuous support of the Submicron Center staff, the support of the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013)/ERC under grant 713351, and the partial support of the Minerva Foundation under grant 713534. M.B. acknowledges the support of SNSF Eccellenza grant PCEGP2_194528. **Author contributions:** B.D. and M.H. designed the experiment. B.D. fabricated the devices, performed the measurements, and did the analysis. B.D., M.B., and M.H. participated in discussions and presented the data. V.U. developed and grew the heterostructures supporting the 2DEG. All authors contributed to the manuscript writing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All the data presented in this paper are publicly available at Zenodo (*31*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm6571
Materials and Methods
Supplementary Text
Figs. S1 to S7
Reference (*32*)

Submitted 7 November 2021; accepted 28 July 2022
10.1126/science.abm6571

EL NIÑO

El Niño frequency threshold controls coastal biotic communities

Jack M. Broughton[1]*, Brian F. Codding[1], J. Tyler Faith[1,2,3], Kathryn A. Mohlenhoff[4], Ruth Gruhn[5], Joan Brenner-Coltrain[1], Isaac A. Hart[1,6]

El Niño has profound influences on ecosystem dynamics. However, we know little about how it shapes vertebrate faunal community composition over centennial time scales, and this limits our ability to forecast change under projections of future El Niño events. On the basis of correlations between geological records of past El Niño frequency and the species composition of bird and fish remains from a Baja California bone deposit that spans the past 12,000 years, we documented marked faunal restructuring when major El Niño events occurred more than five times per century. This tipping point has implications for the past and future ecology of eastern Pacific coastal environments.

The El Niño/Southern Oscillation (ENSO) is a major source of global climatic variability. It is best known for its "El Niño" phase that is characterized by warm eastern Pacific sea surface temperature (SST), intense winter storm activity, and heavy precipitation. Based primarily on data and observations from ENSO variability over the past ~100 years, this climate system is widely recognized to have profound influences on marine and terrestrial ecosystem dynamics, drought, wildfire, energy and water use, as well as human health and global economies (*1*–*3*). Although of enormous value, instrumental records do not extend beyond the past century and are therefore limited by the modern range of variation. We thus have little knowledge about how El Niño frequency and intensity drive shifts in marine and terrestrial biotic community composition over centennial or millennial time scales. This gap precludes our ability to forecast long-term changes in biotic communities under future El Niño variation scenarios (*4*). We address this here by demonstrating how centennial-scale marine and terrestrial vertebrate community composition in Baja California varied over the past 12,000 years in relation to El Niño frequency variation as reconstructed from a high-elevation South American lake. The analysis reveals an El Niño frequency threshold beyond which faunas undergo substantial community shifts. This record is derived from a mixed archaeological and raptor deposit from the coastal cave site of Abrigo de los Escorpiones (hereafter, Escorpiones).

El Niño events vary widely in their magnitude, intensity, duration, and geographic reach (*2*, *3*), initiating complex biotic community-level and species-specific responses (*5*–*15*) in modern vertebrate faunas from many areas of the eastern Pacific. For example, in the region adjacent to the California Current System, enhanced precipitation and primary productivity associated with El Niño have been causally linked to expanding population densities of many terrestrial bird and mammal species (*5*–*8*). By contrast, El Niño typically has negative impacts on many regional marine fishes (*9*, *10*), marine mammals (*11*, *14*), and seabirds (*12*, *13*). When high SST prevails during El Niño years, phytoplankton and zooplankton concentrations decline (*15*), leading to shifting distributional patterns and depressed populations of many fish taxa and a range of effects on seabirds and marine mammals, including range changes and declines in breeding productivity, reproductive success, and local abundance. In addition, high-energy storms associated with El Niño have strong negative impacts on nearshore kelp forests that harbor diverse vertebrate faunas (*9*, *10*). ENSO is thus a well-known modern driver of biophysical interactions, but its long-term influence on past faunal community composition is not well known outside of coastal Peru (*3*).

The Escorpiones site is located on the northwest coast of Baja California, Mexico. The site is situated beneath a cliff face topped with prominent raptor perches ~100 m from the Pacific coast and contains an ~5-m-deep stratified shell midden mixed with raptor deposits (Fig. 1 and fig. S1) (*16*). Excavation of the site produced numerous artifacts of stone, bone, and shell (fig. S2) and a massive faunal assemblage, with 97 radiocarbon dates on charcoal and bone forming the basis of a robust depositional chronology that reveals a 12,000-year history of human and raptor deposition (table S1). This dating effort is comparable to the highest-resolution (dates/millennium) late

[1]Department of Anthropology and Archaeological Center, University of Utah, Salt Lake City, UT, USA. [2]Natural History Museum of Utah, Salt Lake City, UT, USA. [3]Origins Centre, University of the Witwatersrand, Johannesburg, South Africa. [4]Paleowest, Salt Lake City, UT, USA. [5]Department of Anthropology, University of Alberta, Edmonton, Alberta, Canada. [6]Department of Geography, University of Utah, Salt Lake City, UT, USA.
*Corresponding author. Email: jack.broughton@anthro.utah.edu

Fig. 1. Location and interior view of Escorpiones. The excavated portion is at the base of the crater face (top, center).

Quaternary vertebrate sequences in North America (*17*, *18*).

A total of 18,623 identified marine fish and marine and terrestrial bird specimens representing 132 taxa are the focus of this study (tables S2 to S8 and figs. S3 to S5). To address how past El Niño influenced this fauna, we draw upon the Lake Pallcacocha (Ecuador) sediment core, one of the most widely applied, continuous late Quaternary, eastern Pacific proxy records of paleo-ENSO (*19*). Variation in the thickness of clastic laminae deposited in this glacial lake basin provides a high-resolution record of moderate-to-strong prehistoric, high-precipitation events (in 100-year intervals) linked to El Niño frequency that is based on strong correlations with instrumental and historical records of these events over the past 200 years. This sequence is consistent with other proxy data and modeling results for paleo-ENSO for the northeastern Pacific (*20*, *21*), including previously documented covariation with the abundance of mesic rabbit and hare (Lagomorpha) fauna from Escorpiones (*22*). The Pallcacocha record indicates a near absence of ENSO during the early Holocene [11.0 to 7.0 thousand years ago (ka)], with a substantial increase occurring between ~7.0 to 5.0 ka [Fig. 2 A and fig. S7 (*19*)] that continues through much of the late Holocene before declining from 1.2 ka to the present. From the radiocarbon ages, age-depth models were created [figs. S9 to S11 (*22, 23*)] to assign faunal materials to 100-year units to align them to this proxy for paleo-ENSO.

On the basis of the influence of ENSO on modern birds of the California Current System, we expect that high El Niño frequency in the past negatively affected marine taxa but enhanced the productivity of terrestrial faunas. Among the fishes, we expect the frequency of fish taxa associated with productive kelp forest habitats, or those that have low resiliency to ENSO-based habitat disturbance [i.e., kelp bass (*Paralabrax clathratus*) and rockfish (*Sebastes* spp.)], to have declined relative to nearshore habitat generalists such as surfperch (Embiotocidae). These trends are each clearly apparent in the Escorpiones record when abundance indices for these groups are plotted by the 100-year time bins across the history of the site (Fig. 2, B and C). The inverse relationships between the faunal indices and El Niño frequency suggest that the latter had a substantial influence on past faunas in ways consistent with their ecological characteristics.

To further explore the response of past faunal communities to ENSO, we performed dimensional reduction analyses [detrended correspondence analysis (DCA) (*24*)] for the bird and fish taxonomic abundance data (tables S6 and S7). We used this ordination technique because the primary axes of the DCAs summarize dominant changes in faunal composition, facilitating the interpretation of all identified specimens from the 130+ taxa represented. We investigated how DCA primary axis scores vary with taxa-specific indices using generalized linear models. To quantitatively evaluate how the bird and fish assemblages captured by the DCA primary axis scores vary as a function of El Niño frequency, we constructed multivariate generalized additive mixed models (*25*) and included El Niño frequency as a smoothed term. The models account for the depositing agent (humans or raptors) by including artifact counts (a relative index of human activity) as a covariate, for differences in sample size by weighting each value by the mean normalized faunal sample size per temporal bin, and for temporal autocorrelation by including an autoregressive correlation structure of order 1 (corAR1; figs. S13 and S14 and table S9). This allowed us to isolate the influence of El Niño frequency on faunal composition. The models specify a Poisson distribution with a log-link and quasi-likelihood estimation, with term smooths selected using restricted maximum likelihood. To identify thresholds in the effect of El Niño frequency on community structure, we examined the first derivatives of the model fits.

The DCA axis scores for the bird and fish assemblages are highly correlated with the relative abundance of marine and terrestrial birds (r^2_l = 0.86; P < 0.0001) and the frequency of fishes with low resilience to ENSO (r^2_l = 0.77; P < 0.0001), respectively. It follows that the change observed for the two faunal abundance indices (Fig. 2, B and C) mirror broader changes in faunal composition observed across the sequence (Fig. 2, D and E). The two DCA axis scores show nearly identical variation over the 12.0 thousand year sequence, with much of it occurring before the onset of a high El Niño frequency and variance phase at ~7.0 to 5.0 ka. However, when El Niño enters this phase, faunal composition undergoes substantial reorganization and, unexpectedly, remains relatively stable across most of the remainder of the sequence.

This trend is further clarified by direct comparisons of El Niño frequency with both

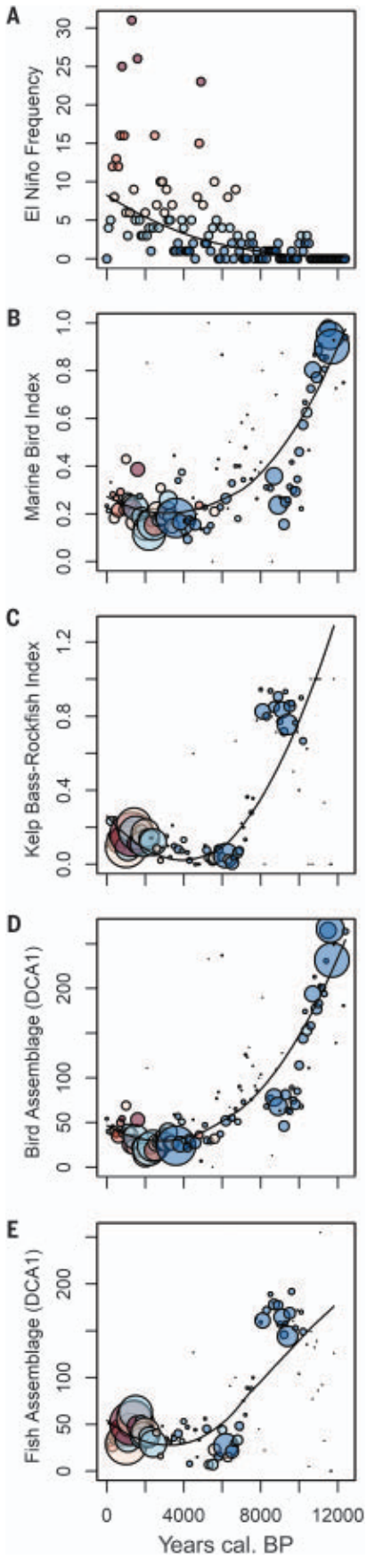


Fig. 2. Trends through time in El Niño frequency (events per century) and faunal composition at Escorpiones fit with a smoothing spline. (**A**) El Niño frequency [from (*19*)]. (**B**) Marine birds = ∑ marine birds/∑ (marine + terrestrial birds). (**C**) Kelp bass-rockfish = ∑ (kelp bass + rockfish/∑ (kelp bass + rockfish + surfperch). (**D**) Primary DCA axis scores for the bird assemblage. (**E**) Primary DCA axis scores for the fish assemblage. Points are colored by El Niño frequency (red, high; blue, low). Point sizes for (B) to (E) correspond to the relative faunal sample sizes.

the bird and fish abundance indices and the DCA axes (Fig. 3). There is extensive temporal variation in faunal composition at low El Niño frequencies, but at higher frequencies, faunal composition is stable through time. Collectively, these data appear to capture assemblage-level, structural changes in the stability and relative productivity of marine versus terrestrial habitats (fig. S12).

This ecological restructuring of the avian and fish communities occurs when five or more moderate-to-strong El Niño events occur during a century (Fig. 3). Thus, assemblages deposited when El Niño events did not exceed this threshold are highly variable in composition, but all of those accumulated at or above this ecological tipping point reflect the relatively stable but low marine- and high-terrestrial productivity state (fig. S12).

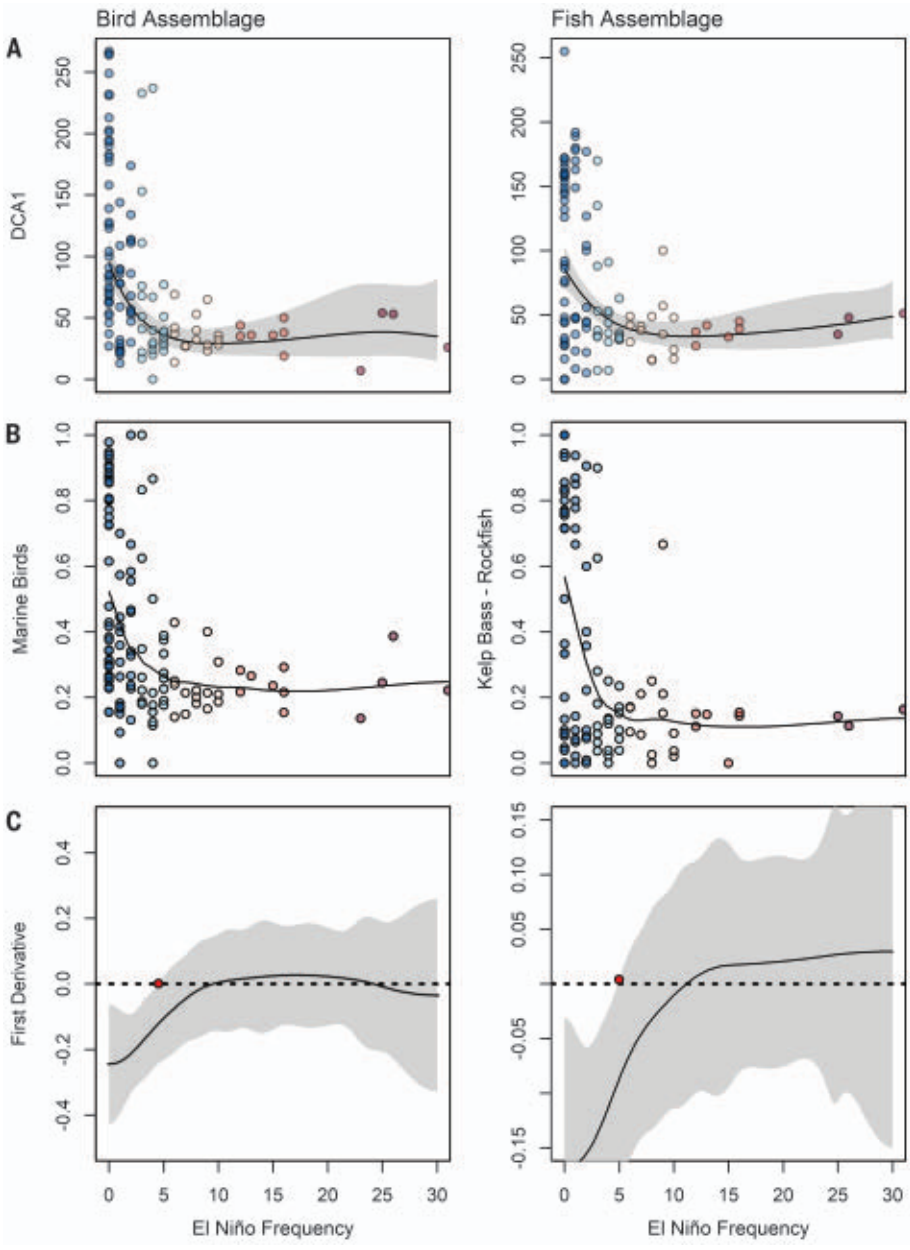


Fig. 3. Primary DCA axes, taxonomic abundance indices, and first derivatives of the model fits for the Escorpiones bird and fish assemblages as a function of El Niño frequency. (**A**) Model fit and 95% confidence intervals shown as a line and gray envelope, respectively. Points are colored by El Niño frequency (red, high; blue, low). (**B**) Relationships between El Niño frequency and the relative abundance of marine avifauna, $\sum$ marine birds/$\sum$ (marine + terrestrial birds), and the frequency of kelp bass and rockfish, $\sum$ kelp bass + rockfish/$\sum$ (kelp bass + rockfish + surfperch). (**C**) Bird and fish DCA–El Niño model fits with 95% confidence intervals showing where the first derivatives are not distinguishable from zero (red circle), which occurs in both cases when El Niño frequency approaches five per 100 years.

During those centuries when El Niño frequencies did not exceed the five-event threshold, which happens throughout the sequence but much more frequently before ~5.0 ka, faunal composition varies widely and is apparently controlled by other climatic or oceanographic drivers [e.g., Pacific Decadal Oscillation (PDO), eustatic sea-level rise, upwelling]. For example, paleo-oceanographic records (e.g., $\delta^{18}O$ and magnesium-calcium ratios from planktonic foraminifera) from the Santa Barbara Basin and Baja California Sur suggest that such factors enabled generally cooler SST and high productivity during much of the early Holocene [fig. S7 (*26*, *27*)].

This ENSO-based faunal community restructuring has implications for the paleoecology, archaeology, and possible future scenarios of eastern Pacific coastal environments. The first concerns the relative use of marine and terrestrial habitats by human foragers. Because El Niño frequency strongly influences faunal productivity and thus foraging returns from marine versus adjacent terrestrial resource patches, we expect greater use of marine habitats for hunter-gatherers during periods with low El Niño frequency. This is the case at Escorpiones, because artifact counts, our measure of human occupational intensity, show that people spent much more time at the site before 7.0 ka, when El Niño frequency did not exceed the five-event-per-century tipping point and other factors converged to produce high marine productivity (Fig. 4A).

This El Niño threshold phenomenon also has implications for our understanding of the human colonization of the Americas because it relates to the productivity of habitats along proposed migratory routes, especially the "Kelp Highway Hypothesis," which suggests that terminal Pleistocene settlers followed a route through stable and productive littoral habitats from Northeast Asia, east across Beringia, and south into the Americas (*28*). Our analysis is consistent with this hypothesis because the faunal communities indicate that high marine productivity occurred throughout the terminal Pleistocene and early Holocene, a period during which the El Niño frequency threshold was not transgressed and other oceanographic processes (e.g., eustatic sea-level change, upwelling, non-ENSO controls on SST) and climate systems (e.g., PDO, North American Monsoon) apparently enabled the formation of productive nearshore habitats in general and kelp forests in particular (Fig. 2A and fig. S12).

The threshold effect also appears linked to the extinction of the goose-sized flightless sea

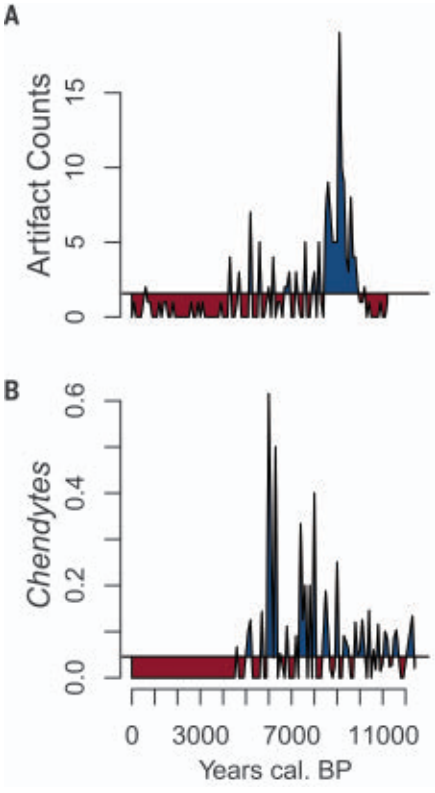


Fig. 4. Time series of artifact counts and *Chendytes* specimens at Escorpiones. (**A**) Artifact counts. (**B**) *Chendytes* = $\sum$ *C. lawi*/$\sum$ marine birds. Blue shading indicates above the time series mean; red shading, values below the time series mean.

duck (*Chendytes lawi*), an issue that has implications for arguments about the human role in past faunal extinctions more broadly. *C. lawi* breeding colonies were confined to offshore rocks and islands off the Pacific coast that provided refugia from mammalian predators, but human foragers were able to access them by watercraft as early as ~10.0 ka (*29*). Archaeological evidence indicates that this duck was harvested until its extinction ~2.5 ka, showing that this vulnerable resource withstood an 8000-year period of human exploitation (*29*).

Escorpiones provides the highest-resolution record of the human use of *C. lawi*, with specimens exhibiting direct evidence of human utilization (stone tool cut marks, burning) from ~10.0 to ~4.0 ka (Fig. 4B and fig. S4). However, its apparent local extirpation at ~4.0 ka occurred just after the ENSO-driven faunal tipping point was reached, driving the shift to lower marine productivity and associated declines in other seabirds (fig. S12). This evidence is inconsistent with the long-debated overkill model for Pleistocene megafaunal extinctions, which requires universal extinctions for vulnerable prey immediately after human arrival (*30*, *31*), but it does not rule out potential human disruption of *C. lawi* demography and habitat and a synergistic climatic-anthropogenic cause for its ultimate extinction.

Ecological thresholds and tipping points caused by environmental drivers are important features influencing the structure and functioning of ecosystems (*32*) and are increasingly used as a framework in environmental management (*33*). Identifying in what ecological contexts and at what temporal-spatial scales such threshold behaviors occur is thus key to deriving practical management implications. Although modern ENSO variability is associated with major interannual faunal shifts, this analysis shows the marine fauna to be more stable but less productive with increasing El Niño events at the centennial scale. This high modern variability is consistent with the low El Niño frequency phase that has characterized the system over the past several centuries (Fig. 2A). We anticipate similar or related threshold features to have characterized this system in vertebrate faunas across its extensive temporal and geographic reach, although identifying them must await the discovery and analysis of additional detailed paleorecords. Here, our analysis indicates that El Niño events have the most controlling effect on biotic assemblages when they occur beyond the critical threshold of five times per century. If strong El Niño events increase in the future, as predicted by some of the latest-generation climate models (*4*), our analysis suggests that ENSO may dampen faunal shifts but move eastern Pacific marine ecosystems back to a less productive state.

REFERENCES AND NOTES

1. K. M. Cobb *et al*., *Science* **339**, 67–70 (2013).
2. A. Timmermann *et al*., *Nature* **559**, 535–545 (2018).
3. D. H. Sandweiss *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8271–8279 (2020).
4. H.-B. Fredriksen, J. Berner, A. C. Subramanian, A. Capotondi, *Geophys. Res. Lett.* **47**, 1–9 (2020).
5. J. H. Brown, M. Ernest, *Bioscience* **52**, 979–987 (2002).
6. J. P. Marshal, P. R. Krausman, V. C. Bleich, W. B. Ballard, J. S. McKeever, *J. Wildl. Manage.* **66**, 1283–1289 (2002).
7. J. A. LaManna, T. L. George, J. F. Saracco, M. P. Nott, D. F. DeSante, *Auk* **129**, 734–743 (2012).
8. M. Holmgren, M. Scheffer, E. Ezcurra, J. R. Gutiérrez, G. M. J. Mohren, *Trends Ecol. Evol.* **16**, 89–94 (2001).
9. A. Bertrand *et al*., "El Niño Southern Oscillation (ENSO) effects on fisheries and aquaculture" (Food and Agriculture Organization of the United Nations, 2020); FAO Fisheries and Aquaculture Technical Paper 660.
10. D. M. Checkley Jr., J. A. Barth, *Prog. Oceanogr.* **83**, 49–64 (2009).
11. B. Shirasago-Germán, E. L. Pérez-Lezama, E. A. Chávez, R. García-Morales, *Estuar. Coast. Shelf Sci.* **154**, 69–76 (2015).
12. D. G. Ainley, R. J. Boekelheide, *Seabirds of the Farallon Islands: Ecology, Dynamics, and Structure of an Upwelling-System Community* (Stanford Univ. Press, 1990).
13. D. Oro, *Front. Ecol. Evol.* **2**, 79 (2014).
14. F. R. Elorriaga-Verplancken, G. E. Sierra-Rodríguez, H. Rosales-Nanduca, K. Acevedo-Whitehouse, J. Sandoval-Sierra, *PLOS ONE* **11**, e0155034 (2016).
15. E. Gómez-Ocampo, G. Gaxiola-Castro, R. Durazo, E. Beier, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **151**, 64–76 (2018).
16. R. Gruhn, A. Bryan, *Pac. Coast Archaeol. Soc. Q.* **42**, 1–16 (2009).
17. R. C. Terry, M. Novak, *Geology* **43**, 487–490 (2015).
18. D. L. Jenkins *et al*., "Geochronology, archaeological context, and DNA at the Paisley Caves," in *Paleoamerican Odyssey*, K. E. Graf, C. V. Ketron, M. R. Waters, Eds. (Texas A&M Univ. Press, 2014) pp. 485–510.
19. C. M. Moy, G. O. Seltzer, D. T. Rodbell, D. M. Anderson, *Nature* **420**, 162–165 (2002).
20. S. M. White, A. C. Ravelo, P. J. Polissar, *Geophys. Res. Lett.* **45**, 316–326 (2018).
21. Z. Liu *et al*., *Nature* **515**, 550–553 (2014).
22. I. A. Hart, J. M. Broughton, R. Gruhn, *Quat. Res.* **84**, 46–56 (2015).
23. M. Blaauw, J. A. Christen, *Bayesian Anal.* **6**, 457–474 (2011).
24. A. Hammer, D. A. T. Harper, P. D. Ryan, *Palaeontol. Electronica* **4**, 9https://palaeo-electronica.org/2001_1/past/issue1_01.htm (2001).
25. S. N. Wood, *Generalized Additive Models: An Introduction with R* (Chapman and Hall/CRC, ed. 2, 2017).
26. D. J. Kennett, J. P. Kennett, J. M. Erlandson, K. G. Cannariato, *Quat. Sci. Rev.* **26**, 351–367 (2007).
27. T. M. Marchitto, R. Muscheler, J. D. Ortiz, J. D. Carriquiry, A. van Geen, *Science* **330**, 1378–1381 (2010).
28. J. M. Erlandson *et al*., *Science* **331**, 1181–1185 (2011).
29. T. L. Jones *et al*., *Quat. Sci. Rev.* **260**, 106914 (2021).
30. P. S. Martin, in *Pleistocene Extinctions: The Search for a Cause*, P. S. Martin, H. E. J. Wright, Eds. (Yale Univ. Press, 1967), pp. 75–120.
31. D. J. Meltzer, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28555–28563 (2020).
32. T. M. Lenton, *Annu. Rev. Environ. Resour.* **38**, 1–29 (2013).
33. P. M. Groffman *et al*., *Ecosystems* **9**, 1–13 (2006).

ACKNOWLEDGMENTS

We thank C. Wood and R. Faucett of the Burke Museum, C. Cicero of the Museum of Vertebrate Zoology, and M. Flannery of the California Academy of Sciences for providing access to skeletal reference material; K. Gobelet for providing reference specimens and advice to K.A.M. in her analysis of the fish materials; and the three anonymous reviewers who provided thoughtful comments on the manuscript. **Funding:** This work was supported by the National Science Foundation, Division of Behavioral and Cognitive Sciences, Archaeology (grant 1624189 to J.M.B. and J.C.B.) and the University of Alberta (R.G.). **Author contributions:** Conceptualization: J.M.B., B.F.C., J.T.F., K.A.M., J.B.C., I.A.H.; Formal analysis: J.M.B., B.F.C., J.T.F., J.B.C., R.G., K.A.M., I.A.H.; Funding acquisition: J.M.B., J.B.C., R.G.; Investigation: J.M.B., B.F.C., J.T.F., J.B.C., R.G., K.A.M., I.A.H.; Methodology: J.M.B., B.F.C., J.T.F., J.B.C., R.G., K.A.M., I.A.H.; Project administration: J.M.B., R.G.; Visualization: J.M.B., B.F.C., J.T.F., J.B.C., R.G., I.A.H.; Writing – original draft: J.M.B.; Writing – review and editing: J.M.B., B.F.C., J.T.F., J.B.C., I.A.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm1033
Materials and Methods
Tables S1 to S9
Figs. S1 to S15
Appendix S1
References (*34*–*128*)
MDAR Reproducibility Checklist

Submitted 25 August 2021; accepted 14 July 2022
10.1126/science.abm1033

EVOLUTIONARY BIOLOGY

Evolutionary gain and loss of a pathological immune response to parasitism

Jesse N. Weber[1]*†‡, Natalie C. Steinel[1]*‡§, Foen Peng[2]‡¶, Kum Chuan Shim[1], Brian K. Lohman[1]#, Lauren E. Fuess[2]**, Swapna Subramanian[2], Stephen P. De Lisle[2]††, Daniel I. Bolnick[1,2]*

Parasites impose fitness costs on their hosts. Biologists often assume that natural selection favors infection-resistant hosts. Yet, when the immune response itself is costly, theory suggests that selection may sometimes favor loss of resistance, which may result in alternative stable states where some populations are resistant and others are tolerant. Intraspecific variation in immune costs is rarely surveyed in a manner that tests evolutionary patterns, and there are few examples of adaptive loss of resistance. Here, we show that when marine threespine stickleback colonized freshwater lakes, they gained resistance to the freshwater-associated cestode *Schistocephalus solidus*. Extensive peritoneal fibrosis and inflammation are a commonly observed phenotype that contributes to suppression of cestode growth and viability but also imposes a substantial cost on fecundity. Combining genetic mapping and population genomics, we find that opposing selection generates immune system differences between tolerant and resistant populations, consistent with divergent optimization.

Parasites impose strong selection on their hosts, driving rapid immune evolution (*1*, *2*). However, selection may not always favor the evolution of ever-greater resistance. Resistance-based immunity can have costly side effects that reduce host fitness by consuming limited resources or inducing autoimmune pathology, both of which influence survival and reproduction (*3*, *4*). For example, adult Soay sheep with higher self-reactive antibody titers had lower reproductive rates (*5*), although immune costs have not been found in some other systems, perhaps because costs are condition dependent (*6*). Evolutionary optimization theory suggests that the marginal benefit of evolving increased parasite resistance must be balanced against the marginal costs imposed by the immune response (*7*). As a result, populations may evolve toward an intermediate optimum with partial resistance and some autoimmune costs. Or, populations may evolve tolerance (*8*), allowing infection to persist while mitigating mortality and fecundity costs of that infection (including direct effects of the parasite, or immune costs). Tolerance itself may be costly, either requiring energy for damage repair or by permitting infection. Either resistance or tolerance should fix in a population when these strategies are controlled by separate mechanisms with linear and additive costs; mixed strategies can be maintained under different cost functions (*9*). Here, we show that a particular immune response against parasitic cestodes entails benefits and costs and has been repeatedly gained and lost in a small fish. These results provide a stark example of conflicting selection pressures acting on immune variation in the wild, leading to among-population divergence in immune or tolerance strategies.

Marine populations of threespine stickleback (*Gasterosteus aculeatus*) rarely encounter the freshwater cestode *Schistocephalus solidus* (Fig. 1A) and so have not evolved resistance (*10*). After post-Pleistocene glacial retreat, marine stickleback established populations in freshwater lakes where they repeatedly evolved greater resistance (*10*). However, infection prevalence varies among lakes (Fig. 1A). Ecological factors (i.e., fish size, diet, and lake elevation and area) partly explain infection variation (figs. S1 and S2), but even similar lakes differ in *S. solidus* prevalence. For instance, stickleback from Roberts and Gosling

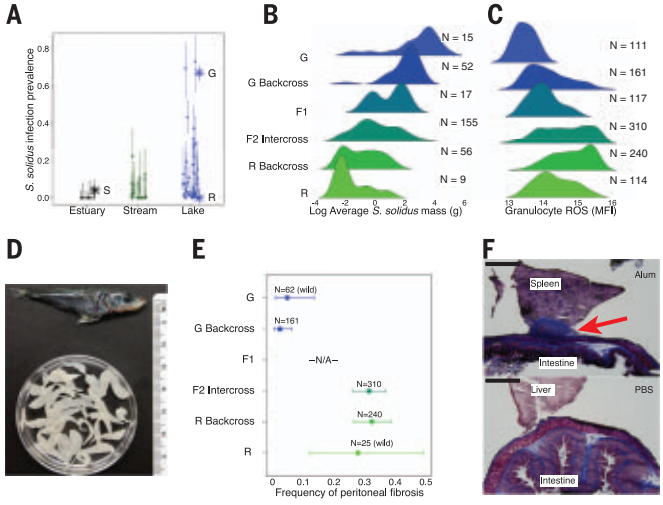


Fig. 1. Population differences in infection success and immune phenotypes. (**A**) *S. solidus* prevalence varies between stickleback in different aquatic habitats on Vancouver Island, British Columbia [binomial GLM deviance = 148.8, degrees of freedom (df) = 2, $P < 0.0001$, 2009 sample]. Prevalence also varies significantly among lakes ($\chi^2 = 4629$, df = 49, $P < 0.0001$; estimate and ±SEM). Three focal populations are labeled with an asterisk: high-infection Gosling Lake (G), infection-free Roberts Lake (R), and a low-infection susceptible marine population from Sayward Estuary (S) representing the ancestral state. (**B**) Cestode mass is largest in lab-infected G stickleback, smallest in R, and intermediate in F1, F2, and reciprocal backcross hybrids. In a general linear model, there was a significant effect of R ancestry proportion ($t = -11.6$, $P < 0.0001$), controlling for covariates including sex ($t = -2.5$, $P = 0.013$), log fish mass ($t = 3.9$, $P = 0.001$), and coinfection intensity ($t = -1.0$, $P = 0.308$). (**C**) Reactive oxygen species production by granulocytes [median fluorescent intensity (MFI) of gated granulocytes] differs between lab-raised G and R stickleback and their F1 and F2 hybrids [sample sizes listed for each cross type; MFI depends on cross type ($F = 43.14$, $P < 0.0001$) and log mass ($F = 27.06$, $P < 0.0001$) but not fish sex ($F = 0.815$, $P = 0.4428$) or cestode presence ($F = 2.06$, $P = 0.1518$)]. A weak cross × infection interaction ($F = 2.47$, $P = 0.0308$) exists because RBC backcross fish show a significant decrease in ROS production when infected, whereas all other crosses exhibit no response. (**D**) Image of infected fish with dissected cestodes. (**E**) Peritoneal fibrosis is more common in wild-caught R than in G fish and is more common in lab-reared cestode-exposed RBC and F2 hybrids than in GBC hybrids (cross $\chi^2 = 155$, $P < 0.0001$; sex $\chi^2 = 2.53$, $P = 0.281$; log mass $\chi^2 = 23.83$, $P < 0.0001$; cestode infection $\chi^2 = 98.48$, $P < 0.0001$; cross × infection $\chi^2 = 18.45$, $P = 0.002$; points are means ± SEM). (**F**) Masson's trichrome stain section of stickleback visceral organs from alum-injected and phosphate-buffered saline control–injected fish. Red arrow indicates fibrosis (blue-stained collagen) connecting the intestinal wall to the spleen. Scale bars, 500 μm.

[1]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA. [2]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA.
***Corresponding author. Email: jnweber2@wisc.edu (J.N.W.); natalie_steinel@uml.edu (N.C.S.); daniel.bolnick@uconn.edu (D.I.B.)**
†Present address: Department of Integrative Biology, University of Wisconsin–Madison, Madison, WI 53706, USA. ‡These authors contributed equally to this work. §Present address: Department of Biological Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA. ¶Present address: Department of Biology, Haverford College, Haverford, PA 19041, USA. #Present address: Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA. **Present address: Department of Biology, Texas State University, San Marcos, TX 78666, USA. ††Present address: Evolutionary Ecology Unit, Department of Biology, Lund University, 223 62 Lund, Sweden.

lakes (hereafter, R and G) on Vancouver Island, British Columbia, exhibit infection rates of 0 and >50%, respectively, despite close geographic proximity, similar consumption of cyclopoid copepods (the first host) (*11*), and nesting pairs of loons in both lakes (the final host). We hypothesized that these differences in infection prevalence reflect either evolution of resistance by R stickleback or tolerance of G fish.

We experimentally exposed lab-raised R×G recombinant hybrids, including intercross (F2) and backcross hybrid stickleback, to *S. solidus* (fig. S3) (*12*). Infection rates were lower in hybrids with greater R ancestry (fig. S4), consistent with prior trends (*10*, *11*). R ancestry also reduced the growth rate of cestodes that established infections (Fig. 1B). Because all parasites grew for the same period of time (*12*), growth rate is proportional to ln(cestode mass). When the cestode reaches reproductive size (50 mg), it manipulates fish to facilitate predation by birds (terminal hosts) (*13*). Cestode growth suppression in R-backcross (RBC) and many F2 stickleback prevented the parasite from reaching this size threshold, thus protecting the fish from increased predation risk.

Infection increased head kidney granulocyte abundance in both fish genotypes (fig. S5), However, R ancestry increased the inflammatory reactive oxygen species (ROS) generated per granulocyte (Fig. 1C). R genotypes also exhibited a visually distinctive response to infection involving severe peritoneal fibrosis (*14*) (Fig. 1D). In vertebrates, tissue damage from parasites can cause inflammation, fibroblast proliferation, and collagen deposition (fibrosis) forming adhesions between the viscera. In healthy stickleback, the abdominal organs move freely, but cestode exposure induces extensive adhesion in some stickleback. Similar adhesions can be induced by intraperitoneal injection of aluminum phosphate (alum) or cestode protein (*15*) and have been observed in a wide variety of other fish species (*16*). Trichrome stain confirms that the adhesions entail excessive collagen deposition (Fig. 1E and fig. S6). After cestode exposure of F2 hybrid fish, fibrosis was positively associated with R ancestry (Fig. 1D), especially when a cestode was present (fig. S7). Fibrosis is rare in G fish (both in the wild and in the lab) and in G-backcross (GBC), regardless of infection. Marine stickleback (a proxy for the ancestral state) also lack fibrosis in the wild and after lab infections. Fibrosis is common in independently colonized watersheds in both Alaska and Vancouver Island, confirming that fibrosis evolved repeatedly during freshwater colonization (fig. S8).

Fibrosis confers two benefits to lab fish. First, fibrotic F2 hybrids had 87.7% smaller cestodes than did nonfibrotic F2 (Fig. 2A; mean mass: 0.316 and 2.558 g, respectively). Considering all infected hybrids, cestode mass depended on fibrosis ($P = 0.0072$) and cross direction ($P < 0.0001$), but did not depend on a fibrosis × cross interaction. ROS production was also negatively correlated with cestode growth within F2 hybrids (Fig. 2B). Second, fibrosis increased the likelihood that small cestodes would be encased in cysts [fig. S9; RBC odds ratio (OR) = 6.6, $P < 0.0001$; F2 fish OR = 13.2, $P < 0.0001$; GBC fish OR = 52, $P = 0.0359$], as collagen is a key component in cyst formation. These cysts often contained dead or visibly degraded parasites, or amorphous material with cestode RNA. Cysts were more common in R ancestry fish (Fig. 2C). Thus, fibrosis is associated with both growth suppression and elimination of cestodes. Fibrosis is also a persistent lesion that remains long after the immune challenge (*15*), which is why we observe fibrosis in cestode-exposed but uninfected fish. The negative effect of R ancestry on cestode growth arises via multiple direct and indirect paths, explaining 34.3% of cestode size variation (Fig. 2D): R ancestry significantly promotes fibrosis and ROS, both these responses independently contribute to cestode growth suppression, and R ancestry explains additional variation in cestode size.

These laboratory results are corroborated by data from wild stickleback populations. In a 2016 sample of 13 lakes in British Columbia, fibrosis prevalence varied significantly among lakes ($\chi^2 = 90.8$, $P < 0.0001$). Fibrosis is more likely in fish with cestode infections (fig. S10) but is unrelated to the only other common helminth in the peritoneum (*Eustrongylides* sp., $\chi^2 = 0.98$, $P = 0.197$). Fibrosis also drove a 56.6% reduction in cestode size (Fig. 2E), controlling for host mass and lake. A 2013 survey of 16 other lakes found that population fibrosis frequency was positively related to cestode prevalence [correlation coefficient (r) = 0.580] and negatively related to mean cestode mass (Fig. 2F).

Both fibrosis and ROS are costly responses to infection. In our experimentally exposed hybrid lab fish, fibrotic females were 73.4% less likely to be gravid at euthanasia [Fig. 2G; binomial general linear model (GLM) $P = 0.0002$]. Additionally, reproductive readiness was negatively related to ROS (fig. S11), increased with fish mass, and was unrelated to cross or infection. These results indicate that cestode infection undermined female fitness indirectly via fibrosis, rather than directly. Field samples corroborate this inference: In 2018, Roselle Lake infected fish were more likely to exhibit fibrosis. Female reproductive maturity was reduced 90.5% by fibrosis (Fig. 2H) but was not significantly affected by infection itself. Similarly, male stickleback with fibrosis were less likely to nest (OR = 0.41 in Roselle Lake, OR = 0.56 in Boot Lake, both $P < 0.01$), controlling for infection status and body size (*17*).

Given these benefits and costs of fibrosis, we hypothesized that natural selection may be optimizing fibrosis, favoring gain (or loss) depending on a population's infection risk or ecological factors affecting costs. Optimization should be reflected in the genetic architecture of the relevant traits; rather than systematic directional selection for ever-stronger fibrosis, stabilizing selection would act on a combination of pro- and antifibrotic genes (positive and negative effects). To test these optimization expectations, we used a triangulation approach merging quantitative trait locus (QTL) mapping, population genomics, and transcriptomics. We mapped the genetic basis of variation in fibrosis (and related traits) and determined QTL effect directions and which populations experienced selection for gain or loss of function.

To map infection-related traits, we genotyped 647 of the experimentally exposed F2, GBC, and RBC fish for 234 informative markers. Mapping QTLs for all traits of interest (fig. S3) revealed seven significant QTLs explaining variation in cestode mass, fibrosis, granulomas, and ROS but not cestode presence or absence (Fig. 3A, Table 1, table S1, and figs. S12 and S13). Fibrosis maps to a QTL on chromosome 2

Table 1. The locations of quantitative trait loci contributing to variation in cestode mass and stickleback immune traits.

Chromosome	Marker ID	Locus	Genomic position	Trait
4	X86	2.3	1406842	Cestode mass
12	X308	17.8	14316110	Cestode mass
15	X359	25.4	1930615	Cestode mass
2	X52	7.8	6190955	Fibrosis
3	X78	34	11999766	Granuloma
11	X275	46.3	14553738	ROS
15	X362	47.7	4462742	ROS

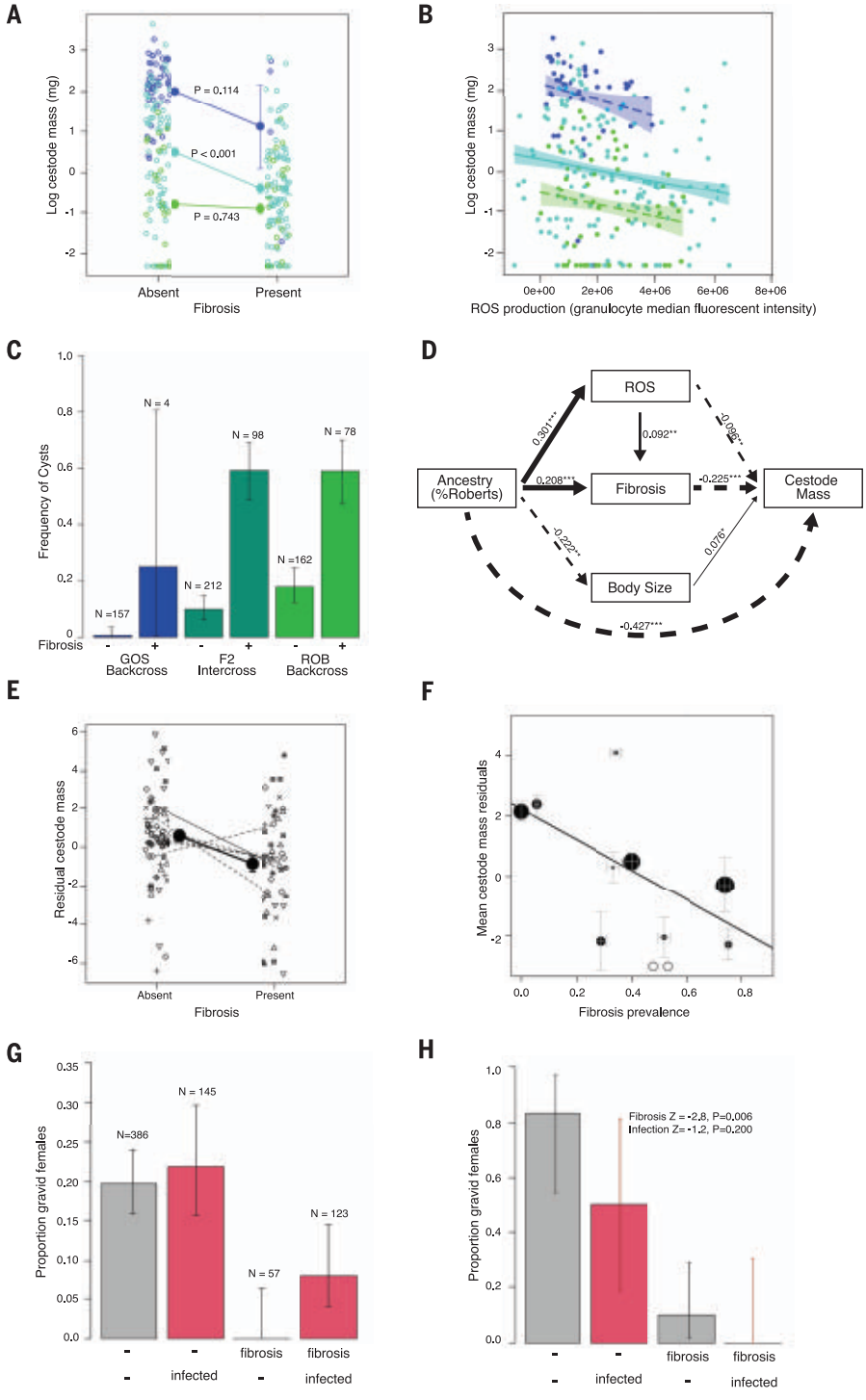


Fig. 2. Benefits and costs of fibrosis and ROS in lab and wild stickleback. (**A**) Cestode mass is lower in fish with greater R ancestry (t = −11.6, P < 0.0001; sex t = −2.505, P = 0.0127; log mass t = 3.86, P = 0.0014; number of coinfecting cestodes t = −1.02, P = 0.3082) and when fibrosis is present (t = −2.909, P = 0.0039), with no cross × fibrosis interaction ($F_{2,255}$ = 1.25, P = 0.2873). Cross types are color coded green (R backcross), turquoise (F2 intercross), and blue (G backcross). (**B**) Stickleback with greater ROS production have smaller cestode mass ($F_{1,236}$ = 5.25, P = 0.0228) controlling for fibrosis ($F_{1,236}$ = 13.18, P < 0.001) and cross ($F_{2,236}$ = 37.9, P < 0.001), with no ROS × cross interaction (P > 0.05). (**C**) Stickleback with fibrosis are more likely to have cysts encasing tapeworms (pictured in fig. S9), particularly in F2 and RBC fish (R ancestry effect χ^2 = 83.9, P < 0.0001). (**D**) Path analysis to explain variation in cestode mass: dashed lines denote negative correlations, solid lines are positive. Thick lines are significant, thin lines are nonsignificant (*P < 0.05, **P < 0.01, ***P < 0.001). (**E**) Wild-caught individuals with fibrosis tend to have lower cestode mass than fibrosis-free individuals. Each trendline is a population (sampled in 2016); solid and dashed lines represent significant and nonsignificant within-lake trends, respectively. The main effect of fibrosis is significant in a model with random population effects (χ^2 = 20.9, df = 1, P < 0.0001, controlling for host mass χ^2 = 0.075, df = 1, P = 0.7833). (**F**) Wild stickleback populations with more fibrosis tend to have smaller cestodes (t = −1.91, one-tailed P = 0.048, controlling for host mass and infection intensity, 2013 sample), confirming lab results. Circle size represents infection prevalence (t = −1.91, controlling for crowding infection intensity). (**G**) Fibrosis, but not cestode infection, was associated with reduced ovary development in lab-raised fish (20.4 versus 5.4% gravid in nonfibrotic versus fibrotic females, respectively; binomial GLM deviance = 13.53, P = 0.0002). (**H**) Fibrosis was associated with low female fecundity in wild fish from Roselle Lake in 2018: 70% of nonfibrotic females were gravid compared with 6.6% of fibrotic females (z = −2.7, P = 0.0057), with no effect of cestode infection. Error bars are ±1 SEM.

(Chr2), where R alleles increase fibrosis frequency (Fig. 3B). Each QTL explains a small-to-moderate amount of phenotypic variance in resistance traits. For example, a QTL on Chr12 explains 9.3% of cestode mass variation (Table 1). Several traits map to overlapping QTLs (e.g., ROS and cestode mass map to Chr15) (Fig. 3A and Table 1), representing potential pleiotropy. An optimization process is expected to yield a mixture of positive- and negative-effect QTLs (*18*) in a given population. We observe such a mixture of effect directions for some traits (ROS and cestode mass). Although R granulocytes generally produce more ROS, QTLs are not uniformly in this direction; within the ROS QTL on Chr15, R alleles confer higher ROS, but at the QTL on Chr11, R alleles reduce ROS (fig. S14).

QTL mapping identifies large chromosomal regions associated with a phenotype, containing large numbers of genes. Genomic signatures of natural selection are typically much narrower, allowing a more precise genetic map. We obtained whole-genome sequencing of pooled population samples (Pool-seq) for R, G, and a marine population (Sayward, or S) and calculated the population branch

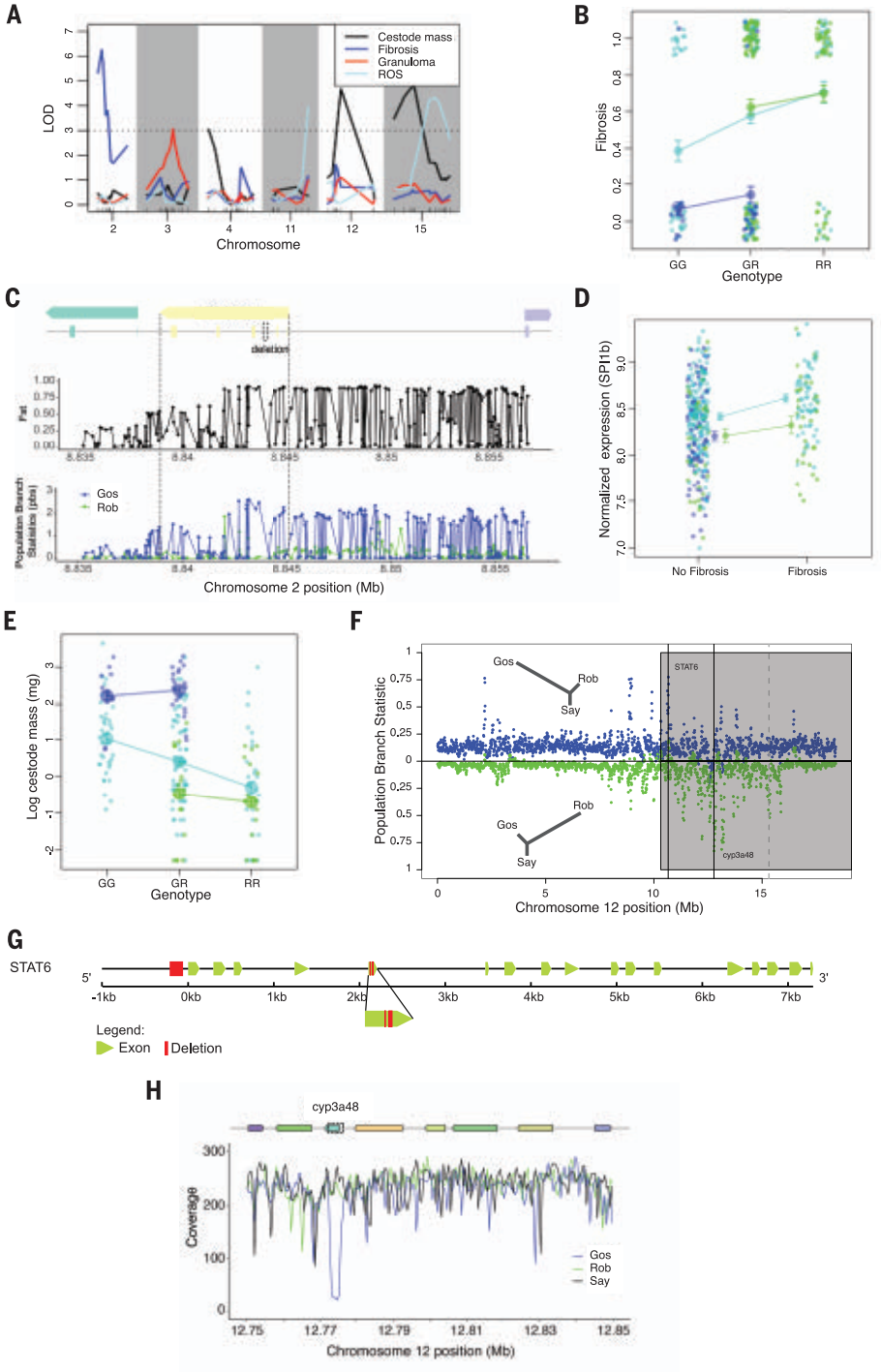


Fig. 3. Genetic mapping and population genomic analyses of immune differences between stickleback populations. (**A**) QTL mapping identified chromosomal regions associated with R versus G differences in cestode growth, fibrosis, granuloma (without fibrosis), and ROS; here, we present logarithm of the odds (LOD) scores for focal QTL. Tick marks on the *x* axis represent population-informative single-nucleotide polymorphisms for mapping. (**B**) Effect plot showing association between Chr2 marker X52 genotype and the frequency of fibrosis, for the three hybrid cross types (blue, RBC; turquoise, F2 intercross; green, GBC), with means and standard errors. Jitter is added to distinguish overlapping points. (**C**) Within the Chr2 QTL for fibrosis, Pool-seq data indicate that the strongest target of divergent selection between R and G is in and adjacent to the 3′ end of the *PU1* gene [top subpanel is F_{ST} between R and G, bottom subpanel is the PBS for G and R showing accelerated evolution is in Gosling Lake, including the fixation of a deletion within the intron (nucleotides 8843941 to 8844018) containing a regulatory CTCF binding motif]. (**D**) The gene *SPI1b* (ENSGACT00000020522.1) produces PU.1 and is more highly expressed in fibrotic than in nonfibrotic fish, controlling for infection and cross [log fold change (LFC) = 0.199, P_{adj} = 0.0324), and is more strongly expressed in infected fish (LFC = 0.254, P_{adj} = 0.0036). (**E**) Effect plot of Chr12 marker X308 genotype on log live cestode mass. (**F**) Within the Chr12 QTL for cestode mass (shaded gray region), the strongest genomic targets of selection (i.e., highest PBS) are tightly clustered around the genes *STAT6* and *Cyp3a48*. (**G**) *STAT6* contains nearly fixed deletions in Gosling Lake just 3′ to the start of the gene and within exon 5. (**H**) *Cyp3a48* contains a 3-kb deletion within exon 2 that is fixed in G fish, as indicated by the large window of reduced coverage in G but not R or Sayward Pool-seq data. In (B), (D), and (E), error bars are ±1 SEM.

statistic (PBS) to detect targets of selection. We focus on genes that (i) lie within a trait QTL, (ii) exhibit strong evidence of selection, (iii) have a known effect on the QTL trait, and (iv) show some feature of molecular evolution, such as changes in coding or known regulatory regions, or are differentially expressed between R and G genotypes (*19*, *20*). Only a few loci met three or four of these criteria. PBS analysis, using marine fish as the outgroup, allowed us to determine which population experienced selection on a given gene. Several of the loci we identified experienced strong positive selection in the susceptible and tolerant G population, others evolved in the resistant R population.

Within the fibrosis QTL (Chr2; Fig. 3B and fig. S15), the strongest target of selection is a narrow area containing the gene *SPI1b* [Fig. 3C and fig. S16; fixation index (F_{ST}) ~ 1.0]. This gene produces a transcription factor, PU.1, previously shown to regulate fibroblast polarization and tissue fibrosis (*21*, *22*). Pool-seq data reveal a 78–base pair (bp) deletion in the second intron that is nearly fixed in G but absent in R and S (Fig. 3C). The Ensembl human genome browser identifies a CTCF binding site in this region, but sequence identity is poor. *SPI1b* is expressed more in fibrotic than nonfibrotic F2 hybrid fish (Fig. 3D), and more in

infected than uninfected fish. Additional genes within the QTL are differentially expressed between genotypes or as a function of fibrosis (fig. S17). But *SPI1b* is the only gene in the QTL to meet all four criteria, providing a strong candidate gene to explain the fibrotic lesions of R fish. Notably, this gene exhibits a large allele frequency change in Gosling rather than Roberts Lake (Fig. 3C, larger PBS), implying that selection favored evolutionary loss of fibrosis in G fish. This result is corroborated by Pool-seq data from 12 high-fibrosis and 13 low-fibrosis populations, showing that *SPI1b* evolution accelerated in about half of the low-fibrosis populations (fig. S18).

The strongest QTL for cestode mass is on Chr12 (Fig. 3E) and contains two strong targets of selection. One locus shows rapid evolution in G and contains *STAT6*, which regulates inflammation and activation of alternatively activated macrophages (*23*) and helminth resistance in lab mice (*24*). G is nearly fixed for a small frame-shift deletion in exon 5 of *STAT6* (Fig. 3G) that should render mRNA isoforms with this exon nonfunctional. G fish also evolved a 100-bp deletion 80 bp upstream of the *STAT6* start codon. The second selected locus within the QTL is driven by evolution in R fish (Fig. 3F and fig. S19) and contains a cytochrome P450 subfamily 3a gene–*Cyp3a48*. Proteins in this subfamily metabolize a large variety of endogenous and exogenous products, with inflammatory ROS being major enzymatic by-products of these reactions (*25*). G fish are nearly fixed for a 3-kb coding-sequence deletion in *Cyp3a48* (Fig. 3H); deletion of this gene is consistent with G fish's weaker fibrosis response. *Cyp3a48* is also among the strongest targets of selection between each of two benthic and limnetic stickleback species pairs (*26*), which also differ in *S. solidus* prevalence. A third site within the QTL is a weaker target of selection (fig. S19E) but is noteworthy for containing *hnf4α*, a fibrosis suppressor (*27*). Ingenuity Pathway Analysis of transcriptome data identified *hnf4α* as an upstream regulatory switch responsible for differential expression of a set of genes (*20*) that are induced by infection [z-score = 0.481, adjusted P value (P_{adj}) = 0.032] in G but not R fish (z-score = −0.8, $P_{adj} < 0.0001$). Infection-induced expression of this fibrosis suppressor in G fish may contribute to their lack of fibrosis. But it is R that experienced selection on this locus (high PBS), leading us to conclude that R fish evolved fibrosis in part by reducing this down-regulatory pathway. Selected regions within other QTLs are presented in fig. S20.

In evolutionary ecology, it is commonly assumed that immunity is preferable to susceptibility (*28*), and hence population differences in parasite resistance reflect evolutionary gain of function in the more resistant populations [but see (*29*)]. Experiments with birds have confirmed that greater resistance can evolve despite short-term immune costs (*30*). However, cost-benefit trade-offs mean that selection may favor either loss or gain of resistance, as we illustrate here. Although marine stickleback have the ancient genetic pathway for peritoneal fibrosis (*16*), this response is not activated after cestode exposure, leading to large, virulent parasites. Replicated evolutionary gain of a strong fibrosis response to *S. solidus* (fig. S8) across numerous freshwater stickleback populations indicates that fibrosis is an adaptive trait. The adaptive benefits of fibrosis include suppressed cestode growth and increased cyst formation and cestode death. These benefits are noteworthy because fibrosis in humans is seen mostly as a pathology (*31*). In stickleback, fibrosis has pathological aspects as well, reducing reproductive success in both sexes. Given this cost-benefit balance, optimization theory leads us to expect stabilizing selection on immune traits leading to both heritable gains and losses of function. Our genetic analyses confirm these expectations. QTL mapping revealed pleiotropic and polygenic evolution of immune traits, with a mix of positive- and negative-effect QTLs (e.g., ROS; fig. S14), consistent with theoretical expectations for stabilizing selection (*18*) rather than relentless directional selection. Using Pool-seq to polarize the direction of evolutionary change within these QTLs, we found strong selection in G favoring rapid fixation of deletions within three profibrotic genes: *SPI1*, *STAT6*, and *Cyp3a48*. We did not detect these deletions in either R or marine fish. Meanwhile, R fish exhibit selection around profibrotic *TMEM39A* and antifibrotic *hnf4α*, whose expression is unresponsive to infection in R fish (but responsive in G fish). Thus, although Roberts Lake fish exhibit a phenotypic gain of function relative to both G and ancestral marine fish, selection actually acted in both lakes in opposing directions (summarized in fig. S21).

Phylogenetic ancestral trait reconstruction also suggests that fibrosis is gained during the repeated colonization of fresh water and then repeatedly lost in a subset of lakes (fig. S8). As a result, cestode resistance and fibrosis severity differ among populations, which contributes to the variation in parasite prevalence among populations. These differences are consistent with theoretical models (*8*, *9*), which suggest that host–parasite coevolution can lead to alternative stable states wherein some populations evolve resistance (despite costs of immunity) and others evolve tolerance (despite costs of infection) (*8*). Such alternative states can allow parasites to persist in host metapopulations, which may facilitate cases of zoonotic disease spillover to humans (*32*). But infection prevalence (and any zoonosis risk) will be unevenly distributed among populations, depending on where they evolve along the resistance–tolerance axis. Thus, to understand parasite epidemiology, we must consider how the mix of immune costs and benefits may lead to evolution toward, or away from, resistance to infection.

REFERENCES AND NOTES

1. M. Fumagalli *et al*., *PLOS Genet.* **7**, e1002355 (2011).
2. D. Ebert, P. D. Fields, *Nat. Rev. Genet.* **21**, 754–768 (2020).
3. M. Nystrand, D. K. Dowling, *BMC Biol.* **18**, 135 (2020).
4. B. C. Sheldon, S. Verhulst, *Trends Ecol. Evol.* **11**, 317–321 (1996).
5. A. L. Graham *et al*., *Science* **330**, 662–665 (2010).
6. K. Norris, M. R. Evans, *Behav. Ecol.* **11**, 19–26 (2000).
7. A. L. Graham, E. C. Schrom 2nd, C. J. E. Metcalf, *Trends Immunol.* **43**, 117–131 (2022).
8. A. Best, A. White, M. Boots, *Evolution* **68**, 1426–1435 (2014).
9. O. Restif, J. C. Koella, *Am. Nat.* **164**, E90–E102 (2004).
10. J. N. Weber *et al*., *Am. Nat.* **189**, 43–57 (2017).
11. J. N. Weber, N. C. Steinel, K. C. Shim, D. I. Bolnick, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6575–6580 (2017).
12. See supplementary materials.
13. J. F. Tierney, F. A. Huntingford, D. W. T. Crompton, *Anim. Behav.* **46**, 603–605 (1993).
14. M. Mittal, M. R. Siddiqui, K. Tran, S. P. Reddy, A. B. Malik, *Antioxid. Redox Signal.* **20**, 1126–1167 (2014).
15. A. K. Hund *et al*., *Evol. Lett.* **6**, 162–177 (2022).
16. M. Vrtílek, D. I. Bolnick, *Evolution* **75**, 2600–2612 (2021).
17. S. P. De Lisle, D. I. Bolnick, *Evolution* **75**, 2509–2523 (2021).
18. H. A. Orr, *Evolution* **52**, 935–949 (1998).
19. B. K. Lohman, N. C. Steinel, J. N. Weber, D. I. Bolnick, *Front. Immunol.* **8**, 1071 (2017).
20. L. E. Fuess *et al*., *Mol. Ecol.* **30**, 6791–6805 (2021).
21. T. Wohlfahrt *et al*., *Nature* **566**, 344–349 (2019).
22. S. Watt *et al*., *Nat. Commun.* **12**, 2298 (2021).
23. Z. Czimmerer *et al*., *Immunity* **48**, 75–90.e6 (2018).
24. J. M. Leung *et al*., *PLOS Biol.* **16**, e2004108 (2018).
25. A. Veith, B. Moorthy, *Curr. Opin. Toxicol.* **7**, 44–51 (2018).
26. A. Härer, D. I. Bolnick, D. J. Rennison, *Mol. Ecol.* **30**, 451–463 (2021).
27. M. M. Yeh, D. E. Bosch, S. S. Daoud, *World J. Gastroenterol.* **25**, 4074–4091 (2019).
28. W. D. Hamilton, M. Zuk, *Science* **218**, 384–387 (1982).
29. L. Råberg, A. L. Graham, A. F. Read, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 37–49 (2009).
30. C. Bonneaud, S. L. Balenger, G. E. Hill, A. F. Russell, *Mol. Ecol.* **21**, 4787–4796 (2012).
31. T. A. Wynn, T. R. Ramalingam, *Nat. Med.* **18**, 1028–1040 (2012).
32. S. Seal, G. Dharmarajan, I. Khan, *eLife* **10**, e68874 (2021).
33. D. Bolnick *et al*., Evolutionary gain and loss of a pathological immune response to parasitism, Dryad (2022); https://doi.org/10.5061/dryad.d51c5b060.

ACKNOWLEDGMENTS

We thank K. Ballare, C. Hernandez, J. Hernandez, M. Kenney, L. Ma, M. Maciejewski, K. Pan, J. Salguero, B. Varela, S. den Haan, and J. Young for research assistance. The Texas Advanced Computing Center (TACC) at the University of Texas at Austin and the Center for Genome Innovation at the University of Connecticut provided HPC resources that contributed to the results reported in this paper. We thank the Genome Sequencing and Analysis Facility of the University of Texas at Austin for sequencing assistance, and the British Columbia Ministry of Environment for permitting fish collection. Images in fig. S21 were created using Biorender.com. **Funding:** This work was funded in part by a Howard Hughes Medical Institute Early Career Scientist fellowship (D.I.B.) and by National Institutes of Health grants 1R01AI123659-01A1 (D.I.B.), 1R01AI146168 (N.C.S.), and 1R35GM142891 (J.N.W.). **Author contributions:** Conceptualization, funding acquisition, methodology, investigation,

analysis, visualization, and writing: J.N.W. and N.C.S. Pool-seq analysis and visualization: F.P. and S.S. Investigation: K.C.S. RNA sequencing (RNA-seq) investigation and analysis: B.K.L. RNA-seq analysis and writing: L.E.F. Field investigation of fibrosis costs and analysis: S.P.D.L. Funding acquisition, conceptualization, methodology, analysis, visualization, writing, editing, and supervision: D.I.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data and analytical code are publicly archived (*33*), and sequence data are available at the Short Read Archive (submission SAMN29177527-SAMN29177550).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo3411

Materials and Methods
Figs. S1 to S21
Tables S1 and S2
References (*34*–*59*)
MDAR Reproducibility Checklist

Submitted 27 January 2022; accepted 21 July 2022
10.1126/science.abo3411

EXOPLANETS

Density, not radius, separates rocky and water-rich small planets orbiting M dwarf stars

Rafael Luque[1,2]* and Enric Pallé[3,4]

Exoplanets smaller than Neptune are common around red dwarf stars (M dwarfs), with those that transit their host star constituting the bulk of known temperate worlds amenable for atmospheric characterization. We analyze the masses and radii of all known small transiting planets around M dwarfs, identifying three populations: rocky, water-rich, and gas-rich. Our results are inconsistent with the previously known bimodal radius distribution arising from atmospheric loss of a hydrogen/helium envelope. Instead, we propose that a density gap separates rocky from water-rich exoplanets. Formation models that include orbital migration can explain the observations: Rocky planets form within the snow line, whereas water-rich worlds form outside it and later migrate inward.

Exoplanets that transit red dwarf stars (M dwarfs) intersect a large fraction of the stellar disk, making them potentially suitable targets for transmission spectroscopy (*1*). The habitable zones of planetary systems around M dwarfs are located close to the host stars, increasing the chance of transits occurring and shortening their period. Whether small planets around M dwarfs are potentially habitable remains unclear, in part because of incomplete knowledge of their composition (*2*).

Small exoplanets are known to have a bimodal radius distribution, with two populations separated by a gap, known as the radius valley (*3*). Potential explanations focus on atmospheric mass loss mechanisms, such as photoevaporation driven by the host star (*4*, *5*) or due to the internal heating of the planet (*6*). Photoevaporation models can reproduce the position of the radius valley by assuming that super-Earth and sub-Neptune planets all have rocky compositions, with their different radii being a consequence of whether they retain their primordial hydrogen/helium atmosphere (H/He envelope). If the internal composition of these planets were icy, the radius valley would be at larger planetary radii (*4*, *7*).

A purely rocky composition for most short-period small exoplanets is inconsistent with global formation models that include accretion and orbital migration mechanisms (*8*). These models predict that planets with masses below 20 Earth masses ($M_\oplus$) become water-rich, because large-planet embryos are preferentially formed beyond the ice line (the distance from the central protostar where the temperature is cold enough for volatile molecules to condense into solid ice grains), and migration dynamics efficiently move objects in this mass range inward (*9*). These models reproduce other observed features of the small exoplanet population, such as the period ratio distribution of adjacent planet pairs and the overabundance of single-transiting systems (*10*). Those results were based on observations of planet radii alone. Knowing the density of each planet might provide more useful information, but this would require measurements of both mass and radius.

We investigated the population of small (planet radius less than 4 Earth radii, $R < 4\,R_\oplus$) transiting planets around M dwarfs (hereafter abbreviated STPMs). Determining the masses of planets observed in transit by space telescopes requires ground-based follow-up, with some STPMs having multiple inconsistent mass estimates in the literature. We compiled all published mass measurements for STPMs and used archival observations to refine the physical parameters of nine planets in seven planetary systems (table S8). We used the JULIET code (*11*) to model transits and radial velocities (*12*). To build our sample of STPMs, we began with the Transiting M-dwarf Planets catalog (*13*), which includes 43 planets with a radius smaller than 4 $R_\oplus$ in 26 planetary systems as of 21 July 2021. We restricted our analysis to planets that are precisely characterized, requiring dynamical mass precision better than 25% and radius precision better than 8%. After our analysis of archival observations (*12*), 34 of the planets (80%) were considered to be precisely characterized by this definition.

Figure 1A compares our STPM sample with theoretical composition models (*14*) on a mass-radius diagram. We find that the planets do not form a continuum but are distributed in three separate populations. Two groups are consistent with specific compositions: the extrapolated mass-radius relation of Earth (hereafter rocky planets), and planets consisting of rock and water ice in 1:1 proportion by mass (hereafter water-rich worlds, or water worlds). The third group consists of planets with larger radii than either model, requiring H/He envelopes. We assigned each planet to the closest model composition, taking into account the uncertainties in mass and radius. In Fig. 1B, we show the same sample in a mass-density diagram, where the bulk densities of the planets have been normalized by a theoretical model of an Earth-like composition (scaled Earth's bulk density, $\rho_{\oplus,s}$, assuming mass fractions of 32.5% iron and 67.5% silicates) that accounts for gravitational compression (*14*).

The rocky population spans a large range of equilibrium surface temperatures (T_{eq}) and has a small dispersion in density. Many of these planets are close enough to their host stars to experience strong runaway greenhouse effects, so they are candidates to have extended atmospheres of water in a supercritical state (*15*, *16*). With little liquid water on the surface, this increases the planetary radius relative to water-free planets (Fig. 2D). The population we identify as water worlds have almost constant bulk densities. These planets must have thin or non-existent H/He atmospheres or supercritical water layers. Otherwise, small variations in the mass fraction of H/He envelopes (Fig. 2A)

[1]Instituto de Astrofísica de Andalucía, Consejo Superior de Investigaciones Científicas, 18008 Granada, Spain. [2]Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA. [3]Instituto de Astrofísica de Canarias, La Laguna, 38205 Tenerife, Spain. [4]Departamento de Astrofísica, Universidad de La Laguna, La Laguna, 38206 Tenerife, Spain.
*Corresponding author. Email: rluque@uchicago.edu

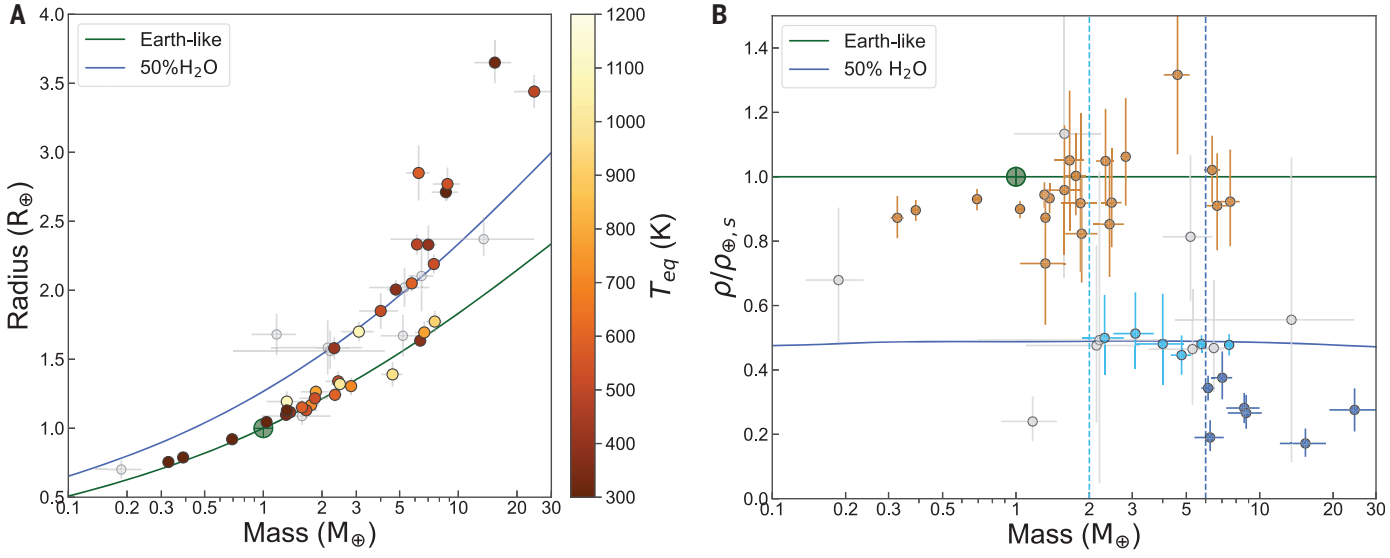


Fig. 1. Sample of small transiting planets around M dwarfs (STPMs) as of 21 July 2021. (**A**) Mass-radius diagram. (**B**) Mass-density diagram. Numerical values are provided in data S1 and include nine planets with revised masses and radii (table S8). Error bars show 1σ uncertainties on each measurement. In both panels, two theoretical composition models (*14*) are plotted: an Earth-like composition (mass fractions of 32.5% iron and 67.5% silicates, green curve) and a planet consisting of 50% water-dominated ices and 50% silicates (blue curve). In (A), planets are color-coded by their equilibrium temperature T_{eq}. In both panels, light gray points are planets with mass or radius determinations worse than our thresholds of 25% and 8%, respectively, so are not included in the subsequent analysis. In (B), densities are normalized by the Earth-like model and planets are color-coded according to their characteristic bulk densities: rocky planets (brown), water worlds (light blue), and puffy sub-Neptunes (dark blue). The vertical dashed lines mark the 2 $M_\oplus$ lower limit for water worlds (light blue) and 6 $M_\oplus$ lower limit for puffy sub-Neptunes (dark blue). For reference, Earth is shown with a green ⊕ symbol.

or in their temperatures (Fig. 2C) would result in large differences in the radius of the planets (*14*–*16*). Therefore, these planets aligned must be water-rich objects, not gas-rich.

The third population have radii larger than 2.3 $R_\oplus$ and masses higher than 6 $M_\oplus$. These are larger than rocky or water-rich planets of the same mass, so we refer to them as puffy sub-Neptunes. The nature of this population is more difficult to determine because interior and atmospheric composition models are degenerate. The possible scenarios include rocky worlds with massive H/He envelopes (Fig. 2A) or water worlds with thin envelopes (Fig. 2B), perhaps affected by a greenhouse effect that generates extended atmospheres of water in a supercritical state (*16*). However, there are no differences in T_{eq} between the water world and puffy sub-Neptune populations, and nearly all planets have T_{eq} > 400 K, high enough to potentially have inflated hydrospheres. Therefore, the larger radii dispersion of puffy planets could be a consequence of the individual H/He accretion histories, not atmospheric loss processes. If so, the water worlds and puffy sub-Neptunes could be part of a continuous population, with differences in their bulk densities arising from their different masses, which affect their accretion potential. Observations of water in the transmission spectra of the puffy sub-Neptunes K2-18 b (*17*, *18*) and HD 106315 c (*19*) are consistent with this scenario, as is the presence of a thick atmosphere that likely contains water in π Mensae c (*20*).

Growth models and Monte Carlo simulations have led to similar conclusions for water worlds (*14*). An issue with this interpretation is that water worlds with masses between 3 and 6 $M_\oplus$ should have radii of 1.5 to 2.0 $R_\oplus$, so no radius valley would be observed (*21*). Figure 1 shows that rocky planets exist from 0.3 to ~10 Earth masses. The radius valley for M dwarfs is not empty: The apparent scarcity of small planets with radii between 1.5 and 2.0 $R_\oplus$ is due to a combination of (i) the rocky population having a maximum mass of 10 $M_\oplus$ whereas the water worlds have a minimum mass of 2 to 3 $M_\oplus$, and (ii) the corresponding radius limits for each (*12*).

Figure 3 shows normalized radius and density histograms for the STPM sample, which we fitted with Gaussian functions. We find that the mean bulk density (radius) in each population is 0.94 ± 0.13 $\rho_{\oplus,s}$ (1.21 ± 0.28 $R_\oplus$) for rocky planets, 0.47 ± 0.05 $\rho_{\oplus,s}$ (1.97 ± 0.28 $R_\oplus$) for water worlds, and 0.24 ± 0.04 $\rho_{\oplus,s}$ (2.85 ± 0.63 $R_\oplus$) for puffy sub-Neptunes. For rocky planets, the average radius is limited by the detection limit of transiting surveys. The separation between super-Earths and sub-Neptunes, which was roughly determined at 1.6 $R_\oplus$ from the radius valley (*22*), is more pronounced in density: We find a clear separation at 0.65 $\rho_{\oplus,s}$ with no overlap between populations.

Figure 4 shows the radius-period and density-period diagrams. We used the GAPFIT (*23*) software to determine the location and slope of the gap that separates rocky planets from water worlds in these two representations. For our STPM sample, we find that the orbital period has no dependence on radius (slope 0.02 ± 0.04). In addition, we find no significant dependence (all slopes are consistent with zero within 1σ) of planet density on orbital period, incident bolometric flux S, or stellar mass $M_\star$ (Fig. 4 and fig. S18).

Our identification of three classes of STPMs according to their bulk density is consistent with formation and evolution theories. Accretion mechanisms predict that ice and rock both participate in planetary growth. Material condensing beyond the water-ice line is expected to have a 1:1 water-to-rock ratio if it has the same composition as the Solar System (*24*), the same ratio we used in Fig. 1. However, the STPM sample has no planets with intermediate water-rock compositions, which are predicted in planetesimal accretion models but not in pebble accretion models (*21*, *25*). Therefore, our classification favors pebble accretion models as the main mechanism for forming small planets around M dwarfs.

Synthetic populations of small planets around M dwarfs (*26*) predict a mass-radius diagram

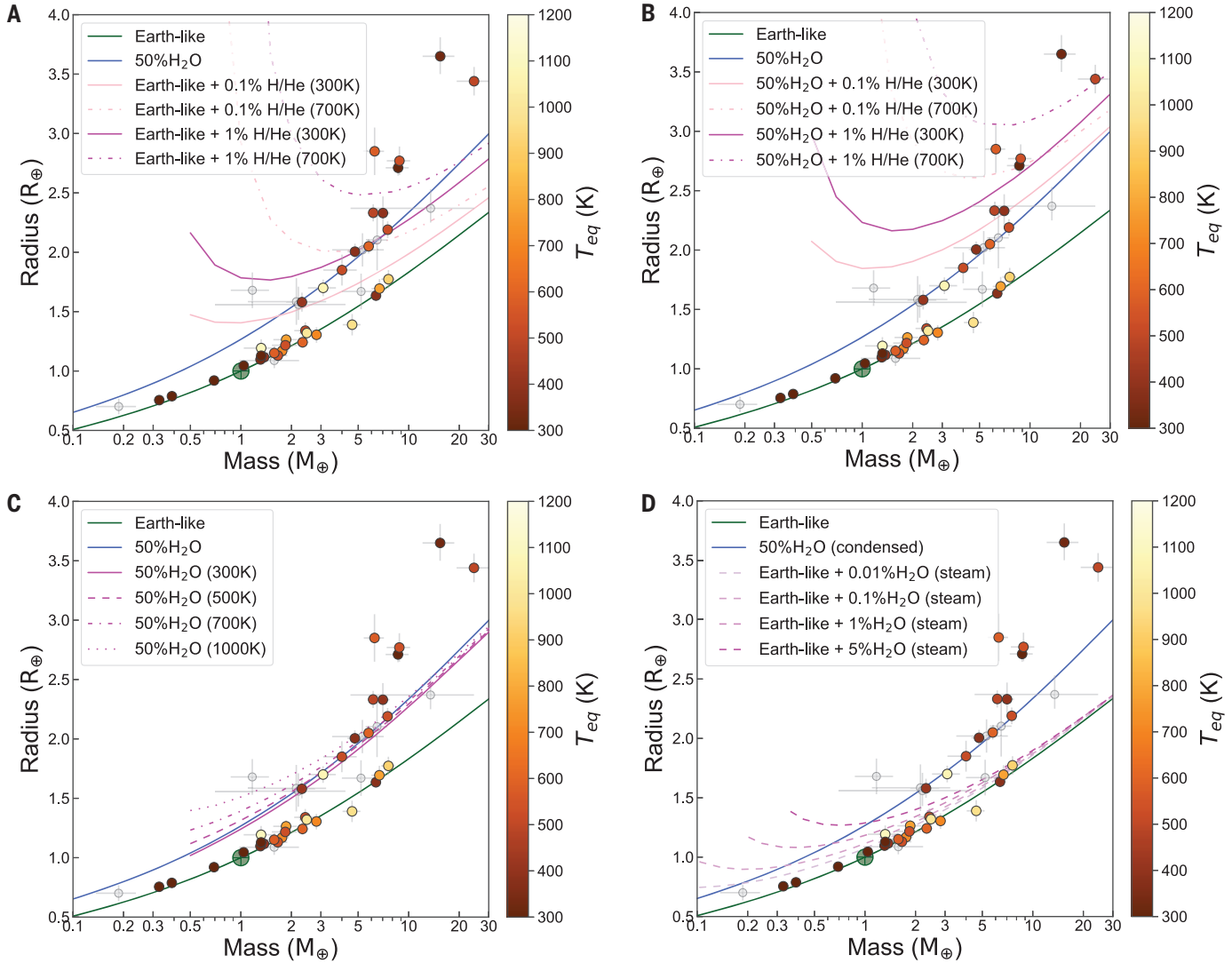
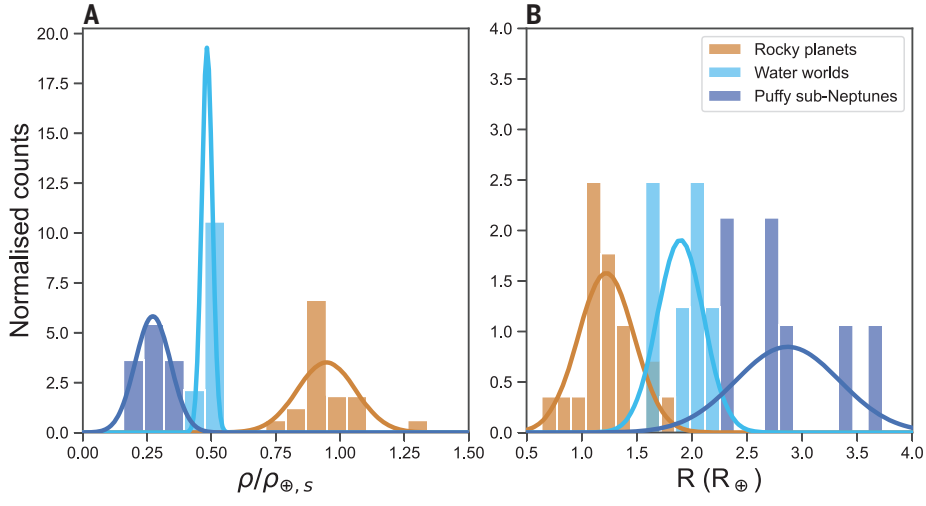


Fig. 2. Same as Fig. 1A, but for different internal composition models. (**A**) Earth-like rocky cores with H/He atmospheres by different percentages in mass at various temperatures (*14*). (**B**) Water-rich cores (50% Earth-like rocky core plus 50% water layer) with different mass fractions of H/He atmospheres at various temperatures (*14*). (**C**) Water worlds at different temperatures (*14*). (**D**) Models for Earth-like planets accounting for runaway greenhouse radius inflation (*15*).

Fig. 3. Normalized histograms of the STPM sample. Only planets that pass our precision requirements are included. (**A**) Frequency as a function of density divided by an Earth-like model. (**B**) Frequency as a function of planetary radius. Colors are as in Fig. 1B. Solid lines show Gaussian models fitted to the distribution of each planet type.



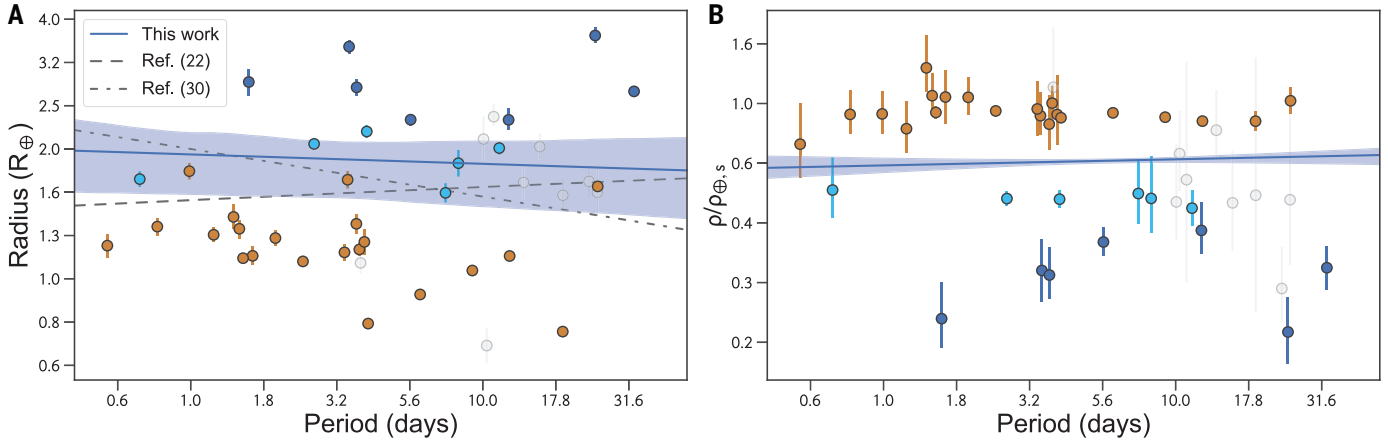


Fig. 4. Dependence of the STPM sample on orbital period. (**A**) As a function of planet radius. (**B**) As a function of planet density. Planets are color-coded according to their bulk density, as in Fig. 1B. The dashed lines are previous determinations of the location of the radius valley (*22*, *30*). The blue line and shaded region show our best-fitting model and its 1σ uncertainty, respectively, which are consistent with zero slope.

that is consistent with our STPM sample (fig. S20). The simulations predict that water worlds are more common at lower stellar masses, with the minimum water world mass being a function of the host star mass. This has been attributed (*26*) to the migration of icy planets from beyond the ice line into the inner regions of the disk. Inward migration becomes efficient at lower planetary masses around lower-mass stars, which do not retain an envelope. For more massive stars, migration only occurs for planets above 10 $M_\oplus$, which are capable of accreting an envelope. Therefore, we propose that the observed population of planets around lower-mass M dwarfs includes more ice-rich cores, with low masses and without envelopes. Our finding of a minimum mass for water worlds of 2 $M_\oplus$ is also in agreement with simulations (*26*) for stellar host masses between 0.3 and 0.5 solar masses, which is the majority of our sample. We conclude that rocky planets formed within the ice line, whereas water worlds (as defined in Fig. 1) formed beyond the ice line and migrated inward. Our sample includes multiplanet systems with planets on either side of the radius valley. For those systems, we find that the innermost planet is always rocky and less massive, whereas the outermost belongs to the water world population (fig. S17).

For solar-type stars, which are higher in mass than M dwarfs, theoretical models predict similar results. Based on mass-radius relations, the planets larger than the radius valley have been identified (*14*) as water worlds, and simulations using global planet formation and evolution models seem to support this hypothesis (*21*). The simulations produce a bimodal distribution of core mass and composition, which agrees with the observations. To explore whether our results can be extended from M dwarfs to solar-type stars, we attempted an analysis of known planets around F-, G-, and K-type stars. The results are shown in fig. S19. The planet distributions share some of the features of the STPM sample; however, the low number of precisely characterized small planets around these stellar types prevents us from drawing conclusions (*12*).

We conclude that STPMs can be classified into three groups using their bulk densities. All three planet types could potentially be habitable if the appropriate conditions are met (*27*–*29*). However, determining those conditions from observations requires knowing the composition of these small planets.

REFERENCES AND NOTES

1. H. Rauer *et al.*, *Astron. Astrophys.* **529**, A8 (2011).
2. A. Segura *et al.*, *Astrobiology* **5**, 706–725 (2005).
3. B. J. Fulton *et al.*, *Astron. J.* **154**, 109 (2017).
4. J. E. Owen, Y. Wu, *Astrophys. J.* **847**, 29 (2017).
5. S. Jin, C. Mordasini, *Astrophys. J.* **853**, 163 (2018).
6. S. Ginzburg, H. E. Schlichting, R. Sari, *Mon. Not. R. Astron. Soc.* **476**, 759–765 (2018).
7. J. G. Rogers, J. E. Owen, *Mon. Not. R. Astron. Soc.* **503**, 1526–1542 (2021).
8. S. N. Raymond, T. Boulet, A. Izidoro, L. Esteves, B. Bitsch, *Mon. Not. R. Astron. Soc.* **479**, L81–L85 (2018).
9. C. Mordasini, Y. Alibert, W. Benz, *Astron. Astrophys.* **501**, 1139–1160 (2009).
10. A. Izidoro *et al.*, *Mon. Not. R. Astron. Soc.* **470**, 1750–1770 (2017).
11. N. Espinoza, D. Kossakowski, R. Brahm, *Mon. Not. R. Astron. Soc.* **490**, 2262–2283 (2019).
12. See supplementary materials.
13. T. Trifonov *et al.*, *Science* **371**, 1038–1041 (2021).
14. L. Zeng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9723–9728 (2019).
15. M. Turbet *et al.*, *Astron. Astrophys.* **638**, A41 (2020).
16. O. Mousis *et al.*, *Astrophys. J. Lett.* **896**, L22 (2020).
17. A. Tsiaras, I. P. Waldmann, G. Tinetti, J. Tennyson, S. N. Yurchenko, *Nat. Astron.* **3**, 1086–1091 (2019).
18. B. Benneke *et al.*, *Astrophys. J. Lett.* **887**, L14 (2019).
19. L. Kreidberg *et al.*, Tentative Evidence for Water Vapor in the Atmosphere of the Neptune-Size Exoplanet HD 106315 c. arXiv 2006.07444 (2020).
20. A. García Muñoz *et al.*, *Astrophys. J. Lett.* **907**, L36 (2021).
21. J. Venturini, O. M. Guilera, J. Haldemann, M. P. Ronco, C. Mordasini, *Astron. Astrophys.* **643**, L1 (2020).
22. R. Cloutier, K. Menou, *Astron. J.* **159**, 211 (2020).
23. R. O. P. Loyd *et al.*, *Astrophys. J.* **890**, 23 (2020).
24. K. Lodders, *Astrophys. J.* **591**, 1220–1247 (2003).
25. N. Brügger, R. Burn, G. A. L. Coleman, Y. Alibert, W. Benz, *Astron. Astrophys.* **640**, A21 (2020).
26. R. Burn *et al.*, *Astron. Astrophys.* **656**, A72 (2021).
27. E. Bolmont *et al.*, *Mon. Not. R. Astron. Soc.* **464**, 3728–3741 (2017).
28. E. S. Kite, E. B. Ford, *Astrophys. J.* **864**, 75 (2018).
29. N. Madhusudhan, A. A. A. Piette, S. Constantinou, *Astrophys. J.* **918**, 1 (2021).
30. V. Van Eylen *et al.*, *Mon. Not. R. Astron. Soc.* **507**, 2154–2173 (2021).

ACKNOWLEDGMENTS

We thank H. Parviainen for help with the photometry analysis of the L 98-59 planetary system. **Funding:** Supported by the Spanish Ministerio de Ciencia e Innovación, through project PID2019-109522GB-C52, the Centre of Excellence "Severo Ochoa" award SEV-2017-0709 to the Instituto de Astrofísica de Andalucía, and the University of La Laguna through the Margarita Salas Fellowship from the Spanish Ministry of Universities ref. UNI/551/2021-May-26 and under EU Next Generation funds (R.L.) and by the Spanish Ministry of Economics and Competitiveness through grant PGC2018-098153-B-C31 (E.P.). **Author contributions:** R.L. analyzed and interpreted the data. E.P. interpreted the data. Both authors wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The Transiting M-dwarf Planets catalog (*13*) is available at https://carmenes.caha.es/ext/tmp/. Our STPM catalog, before and after our revision to the parameters of nine planets, is provided in data S1. The HARPS radial velocity observations we used are available in the ESO Science Archive Facility, http://archive.eso.org/eso/eso_archive_main.html, under target names GJ 1252 and LHS 1815. The TESS observations we used are available in the Mikulski Archive for Space Telescopes, https://archive.stsci.edu/tess/, under target names L 98-59 and LTT 3780. Reduced transit photometry and radial velocity measurements for the systems revised in this work are also provided in data S1. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7164
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S8
Data S1
References (*31*–*94*)

Submitted 30 July 2021; accepted 21 July 2022
10.1126/science.abl7164

METASURFACES

Chiral emission from resonant metasurfaces

Xudong Zhang[1], Yilin Liu[1], Jiecai Han[2], Yuri Kivshar[3]*, Qinghai Song[1,4,5]*

Ultracompact sources of circularly polarized light are important for classical and quantum optical information processing. Conventional approaches for generating chiral emission are restricted to excitation power ranges and fail to provide high-quality radiation with perfect polarization conversion. We used the physics of chiral quasi-bound states in the continuum to demonstrate the efficient and controllable emission of circularly polarized light from resonant metasurfaces. Exploiting intrinsic chirality and giant field enhancement, we revealed how to simultaneously modify and control spectra, radiation patterns, and spin angular momentum of photoluminescence and lasing without any spin injection. The superior characteristics of chiral emission and lasing promise multiple applications in nanophotonics and quantum optics.

One of the main targets of contemporary optics is to efficiently manipulate the polarization state of light emission in an ultracompact form (*1*–*5*). Progress in this field relies on continued breakthroughs in the demonstration of chiral materials and chiral optical cavities (*6*). Chiral materials have been intensively explored, and emission of light with 0.95 degree of polarization (DOP) has recently been demonstrated at room temperature by applying electrical or optical spin injection (*7*, *8*). Research on chiral optical micro- and nanocavities is slightly behind (*6*). By constructing metasurfaces without time-reversal symmetry, a balance between left and right circular polarizations can be broken (*9*–*16*). Consequently, circularly polarized spontaneous emissions with the DOP values up to 0.81 has also been achieved with active nanostructures without spin injection (*16*). However, all conventional approaches suffer from incoherent broadband emission, limited DOP, and large radiating angles. Thus, their practical applications are hindered by low efficiency and energy waste to undesired handedness and emission directions. Although chiral microlasers have been demonstrated, large DOPs appear only in specific power ranges, and their subthreshold performances plummet substantially (*17*). To the best of our knowledge, there are no efficient strategies today to control simultaneously both chiral photoluminescence and chiral lasing in an ultracompact form.

To tackle these challenges, we used the physics of optical bound states in the continuum (BICs), which are peculiar states that can localize light within the radiation continuum for an infinitely long time (*18*–*25*). In physical systems, BICs appear as quasi-BICs with finite but still very large quality (*Q*) factors, allowing the achievement of distinct characteristics of both confinement and radiation of light that have successfully been used to boost performance of nanophotonic devices, including unidirectional emission, ultrafast control, and record-high harmonic generation (*20*, *21*, *25*). Recently, the relationship between chiral fields and quasi-BICs has been uncovered (*21*–*23*), and the construction of quasi-BICs with intrinsic chirality has been suggested theoretically (*26*–*28*), making optical quasi-BICs very attractive for chiral devices.

The main concept underpinning our resonant chiral metasurfaces (Fig. 1A) was created by a square lattice of tilted titanium dioxide (TiO_2) bars placed on a 13-nm indium tin oxide (ITO)–coated substrate. The height of the TiO_2 bars and lattice size are fixed at H = 180 nm and a = 360 nm, respectively. The metasurface is characterized by two deformation (asymmetry) parameters θ (in plane) and α (out of plane), which are defined in Fig. 1B. We used 180-nm polycarbonate (PC) film and 400-nm electron-beam resist [polymethyl methacrylate (PMMA) A2] coated onto the TiO_2 bars. For the case of θ = 0° and α = 0°, C2 symmetries are preserved, and BIC with high Q factor appears at

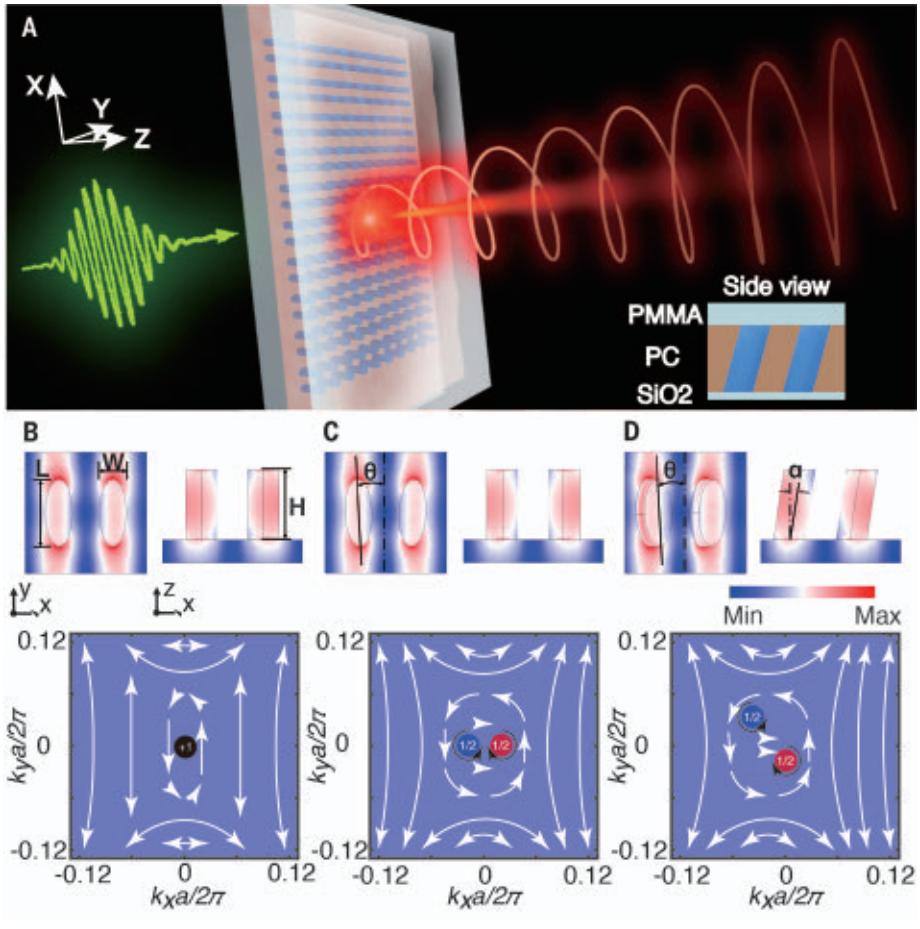


Fig. 1. Resonant chiral metasurfaces. (**A**) Schematic of the nanostructures. (Inset) Orange area indicates DCM-doped PC film. (**B** to **D**) Top view and side view of unit cell with different deformation parameters of (B) θ = 0° and α = 0°, (C) θ = 2.5° and α = 0°, and (D) θ = 2.5° and α = 8.5°. (Bottom) The corresponding directions of polarization vector fields. The length and width of the TiO_2 bar are, respectively, L = 200 nm and W = 80 nm.

[1]Ministry of Industry and Information Technology Key Lab of Micro-Nano Optoelectronic Information System, Harbin Institute of Technology, Shenzhen 518055, P. R. China. [2]National Key Laboratory of Science and Technology on Advanced Composites in Special Environments, Harbin Institute of Technology, Harbin 150080, P. R. China. [3]Nonlinear Physics Center, Research School of Physics, Australian National University, Canberra, ACT 2601, Australia. [4]Pengcheng Laboratory, Shenzhen 518055, P. R. China. [5]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, P. R. China.
*Corresponding author. Email: yuri.kivshar@anu.edu.au (Y.K.); qinghai.song@hit.edu.cn (Q.S.)

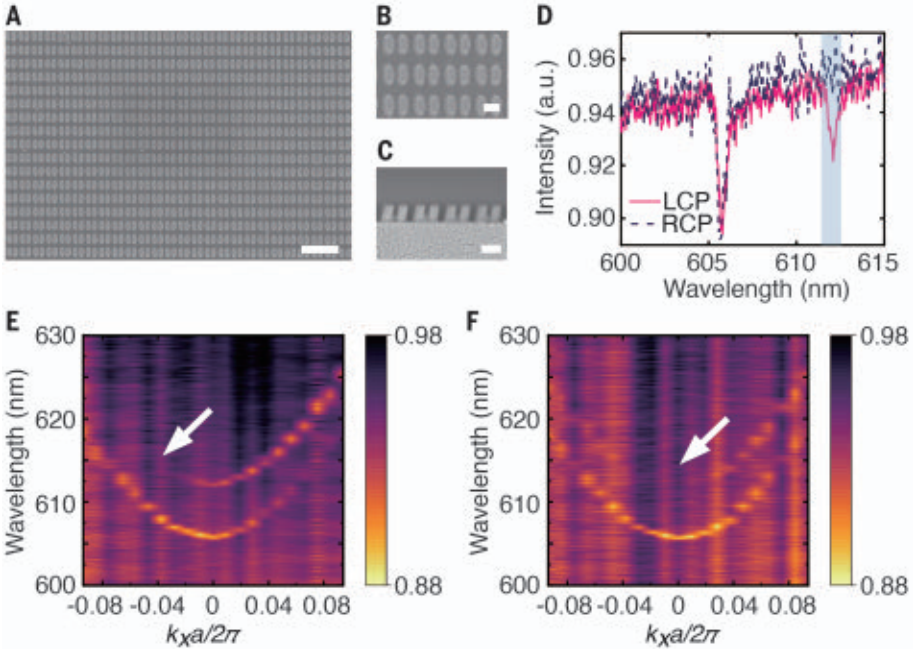


Fig. 2. Experimental demonstration of chiral metasurfaces. (**A**) Top-view SEM image of the sample. (**B** and **C**) The corresponding high-resolution top-view and side-view SEM images. (**D**) LCP and RCP transmission spectra in the normal direction. (**E** and **F**) The angle-dependent LCP and RCP transmission. The C-points are indicated with arrows. The deformation parameters are $\alpha = 8.5°$ and $\theta = 2.5°$. Scale bars, (A) 1 μm; (B) and (C) 200 nm.

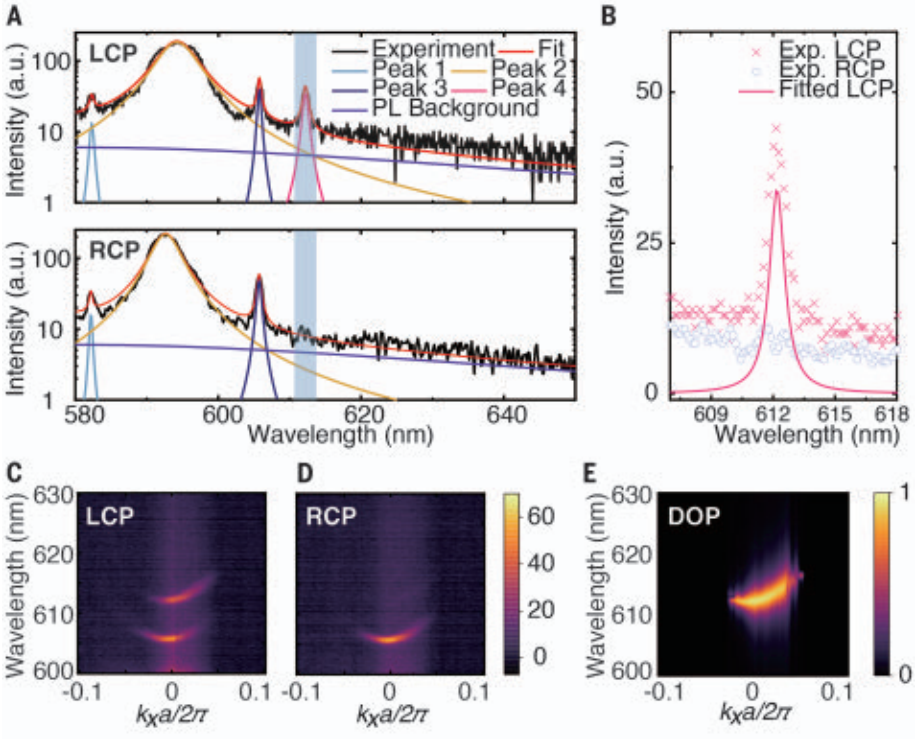


Fig. 3. Observation of chiral emission. (**A**) Spectra of (top) LCP and (bottom) RCP emissions in the normal directions. The fitted lines are shown with different colors. The shaded region is the target mode. (**B**) Fitted LCP component at the designed mode. The raw data (crosses and circles) depict the influence of neighboring modes. (**C** and **D**) LCP and RCP angle-dependent emission spectra from the sample. The scale bars are the same in two subgraphs. (**E**) Angle-dependent DOP after spectral fitting. The pumping power density is fixed at 12 W/cm^2.

the Γ-point (Fig. 1B). With an increase of θ, in-plane C2 symmetry gets broken, and BIC transforms to quasi-BIC with large but finite values of the *Q* factor. Meanwhile, the polarization vortex at the Γ-point splits into two circular polarization points (C-points) with opposite handedness (Fig. 1C) (*21*–*23*). Such C-points relate to external chirality that is strongly restricted in controlling light emission. They deviate from the maximal field enhancement around the Γ-point and lack far-field directionality, resulting in a small DOP and huge waste of emissions (*29*, *30*).

We show that the metasurface performance can be improved when out-of-plane C2 symmetry is further broken by introducing the second asymmetry parameter, the tilting angle α (Fig. 1D). Interaction of field in two bars causes vertical offset and breaks both in-plane and out-of-plane symmetry (*31*). Then, two C-points become asymmetric in the momentum space, and intrinsic chirality appears at the Γ-point ($k_x a/2\pi = 0$, where k_x is x direction wave vector). Numerical simulations show that the tilting angle α can precisely manipulate the positions of C-points in the momentum space (*31*). For $\theta = 2.5°$, a tilting angle of $\alpha = 8.5°$ can move C-points with right-handed circular polarization (RCP) and left-handed circular polarization (LCP) to $k_x a/2\pi \sim 0$ and −0.038, respectively.

The manipulation of C-points can be clearly seen in the transmission spectra, where the RCP and LCP transmission dip of the design mode diminishes at $k_x a/2\pi = 0$ and −0.038, respectively (*31*). The calculated circular dichroism (CD) is as high as 0.994 at two points. Because of the gradual shift from the bottom of the band, the *Q* factor of chiral quasi-BIC reduces from $Q = 40{,}000$ at the Γ-point to $Q = 5000$ at $k_x a/2\pi = -0.038$ (*31*). Thus, even though two generated C-points have similar absolute CD values, their ability to confine light and their impact on emission of light should be totally different (*31*). As a direct comparison, a neighboring mode without precise manipulation of C-points has very close LCP and RCP transmissions, with an intrinsic CD of only −0.08.

The resonant metasurfaces are fabricated with a combined process of electron-beam lithography, lift-off, and reactive ion etching (RIE) (*31*). Our design makes the metasurfaces independent of complicated three-dimensional nanostructures (*26*–*28*) and accessible with a one-step slanted RIE process. The top-view scanning electron microscope (SEM) image of TiO_2 bars is shown in Fig. 2A. The corresponding high-resolution top-view and side-view SEM images are shown in Fig. 2, B and C. The measured structural parameters are length (L) = 195 ± 5 nm, width (W) = 79 ± 5 nm, H =180 ± 5 nm, $\theta = 2.5°$, and $\alpha = 8.5 \pm 0.5°$. The designed metasurfaces were experimentally realized after coating TiO_2 bars with 180-nm

undoped PC (Aladdin) and 400-nm PMMA A2 layers.

Transmission spectra of the metasurface were characterized with a homemade optical setup (*31*). The LCP and RCP transmission spectra in the normal direction are shown in Fig. 2D. For the designed mode, there is one resonant dip in LCP spectrum, whereas the RCP spectrum is almost flat. The full width at half maximum (FWHM) of the chiral quasi-BICs is only 0.49 nm, giving a Q factor of 1250. The corresponding intrinsic CD is around 0.925 ± 0.03 (*31*). The angle-dependent transmission spectra in Fig. 2, E and F, show that the RCP and LCP resonant dips diminish at $k_x a/2\pi$ = 0 and −0.036, confirming the RCP C-point at the Γ-point.

Then, we replaced the undoped PC film with 2-methyl-6-(4-dimethylaminostyryl)-4H-pyran (DCM)–doped PC film (*31*) and measured photoluminescence from the metasurface under excitation of a continuous-wave laser diode with a beam diameter of 15 μm. The LCP and RCP emission spectra in the normal direction are shown in Fig. 3A, top and bottom, respectively [the full spectra is provided in (*31*)]. The fluorescence from the metasurface consists of a series of narrow peaks. Both LCP and RCP emission spectra were well fitted by considering the resonances of the metasurface (*31*). Then, the target mode at 612.08 nm was extracted and plotted in Fig. 3B. It is dominated by LCP components, with negligible RCP emissions. Following the definition of DOP $= \frac{I_{LCP} - I_{RCP}}{I_{LCP} + I_{RCP}}$, the emission at the Γ-point has a DOP of almost 1, which is independent of the spin of the pump laser (*31*).

In the angle-dependent photoluminescence from the metasurface (Fig. 3, C and D), both quasi-BIC mode and its neighboring resonance mostly emit in the normal direction, with divergent angles of 2.5°. Such a divergent angle is more than an order of magnitude narrower than conventional photoluminescence and comparable with that of laser emissions (*20*). Compared with the neighboring mode, the quasi-BIC with its C-point in the vicinity of the Γ-point exhibits an obviously chiral response. It is dominated by LCP fluorescence, whereas RCP emission is at the noise level. The corresponding DOP around the Γ-point is near unity and covers a small angular range of 2.5° (Fig. 3E). These observations are consistent with the above analysis and confirm the capability of chiral metasurface in simultaneous control of spectrum, radiation pattern, and spin angular momentum of light emission.

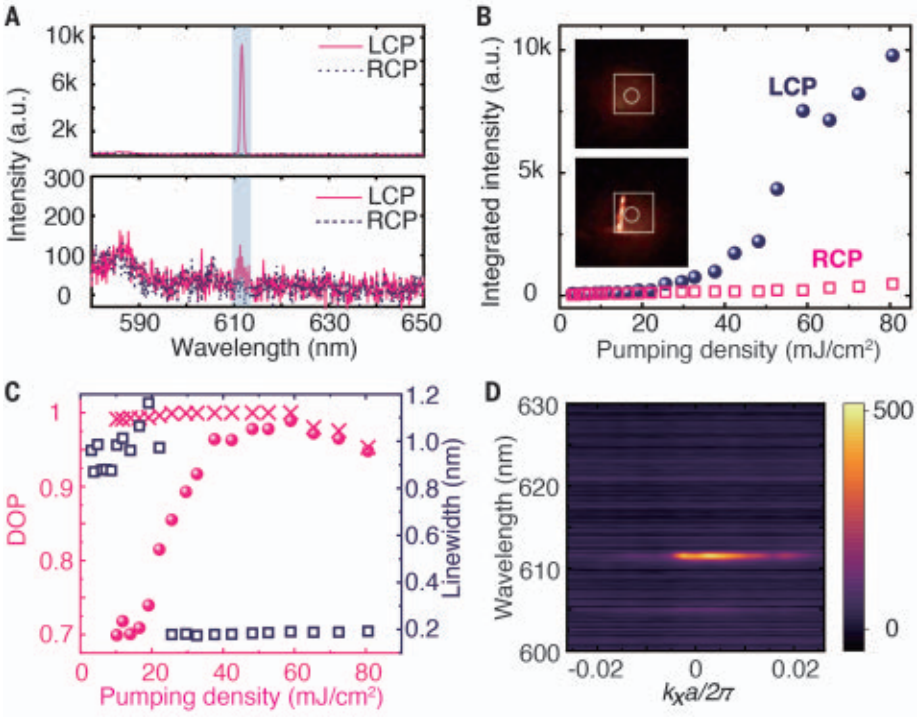


Fig. 4. Observation of chiral lasing. (**A**) LCP (solid lines) and RCP (dashed lines) emission spectra in the normal direction. The pumping densities are 16.5 mJ/cm^2 (bottom) and 58.9 mJ/cm^2 (top), respectively. (**B**) Dependence of output intensity of LCP (dots) and RCP (open squares) on the pumping density. (Insets) The near-field optical image before and after the threshold. The white circles and squares indicate the pumping laser beam and edges of the sample, respectively. (**C**) Absolute DOP (dots), fitted DOP (crosses), and FWHM (squares) of LCP emission as a function of excitation fluence. (**D**) Far-field angular intensity distribution of the chiral microlaser.

The near-unity intrinsic chirality and field enhancement at quasi-BICs are also applicable to ultracompact chiral lasing (*31*). By pumping the samples with a nanosecond laser (beam diameter of 30 μm), light emissions from resonant metasurfaces have been explored. The LCP and RCP emission spectra are depicted in Fig. 4A at different pumping densities. When the pumping density is low, the characteristics of light emission are similar to those in Fig. 3, and the fluorescent microscope image displays a uniform background. At high pumping fluence, the intensity of the LCP mode at 612.08 nm increases dramatically and dominates the emission spectrum. Meanwhile, bright spots appear in the fluorescent microscope image. The dots in Fig. 4B indicate the integrated output intensity of LCP components as a function of pumping density. A dramatic increase of intensity can be seen at the pumping density above 22.14 mJ/cm^2. The open squares in Fig. 4C indicate the power-dependent FWHM of the LCP peak. Associated with the dramatic increase in output intensity, the line width of emission peak reduces from 1.0 to 0.18 nm at the transparent threshold, giving a Q factor of 3400. The dramatic changes in spectrum, FWHM, near-field image, and integrated intensity have confirmed the lasing actions in resonant metasurface.

Compared with spontaneous emission, laser emission at chiral quasi-BICs can have better performances. As stated above, the LCP component experiences an exponential growth above the transparent threshold. The RCP emission looks completely different. Because of the depletion of $|\boldsymbol{E}|^2_{RCP}$ at the Γ-point, where $\boldsymbol{E}$ is electric field in DCM-doped PC, the RCP emission appears as fluorescence and will not reach the laser threshold (Fig. 4B, open squares). Consequently, the difference between LCP and RCP emissions increases dramatically above the threshold. Meanwhile, the gain competition can effectively suppress other modes, with higher thresholds resulting in the single-mode operation. The DOP value, termed as absolute DOP below, will not be affected by the neighboring modes and noises and thus can be directly calculated from the spectra.

The advanced characteristics are shown clearly in Fig. 4A. The absolute DOP of emission from the resonant metasurface above a threshold is as high as 0.989 ± 0.02. The dots in Fig. 4C indicate the absolute DOP values of the chiral microlaser as a function of the pumping density. The absolute DOP value increases rapidly around the transparent threshold saturating at 0.948 to 0.989. This DOP is almost the same as the DOP calculated with spectral fitting (Fig. 4C, crosses). Besides the values of the Q factor and DOP, the far-field directionality of emission is also improved. The gain competition can concentrate almost all the energy to the mode with the highest

Q factor and substantial field enhancement. As a result, the divergent angle is further reduced from ~2.5° to ~1.06° above the threshold (Fig. 4D).

Compared with previous reports (*7*, *8*, *14*, *16*), our resonant metasurface exhibits better performances in terms of DOP, directionality, and Q factor (table S1) that are retained over a wide range, from photoluminescence to lasing (table S2). This approach may improve the design of current sources of chiral light and boost their applications in photonic and quantum systems.

REFERENCES AND NOTES

1. L. Novotny, B. Hecht, *Principles of Nano-Optics* (Cambridge Univ. Press, 2006)
2. Y. J. Zhang, T. Oka, R. Suzuki, J. T. Ye, Y. Iwasa, *Science* **344**, 725–728 (2014).
3. Y.-H. Kim *et al*., *Science* **371**, 1129–1133 (2021).
4. Y. Chen *et al*., *Nat. Rev. Phys.* **4**, 113–124 (2022).
5. A. Forbes, M. de Oliveira, M. R. Dennis, *Nat. Photonics* **15**, 253–262 (2021).
6. H. Hübener *et al*., *Nat. Mater.* **20**, 438–442 (2021).
7. Y. Huang *et al*., *Nat. Photonics* **15**, 475–482 (2021).
8. N. Nishizawa, K. Nishibayashi, H. Munekata, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1783–1788 (2017).
9. Y. Tang, A. E. Cohen, *Science* **332**, 333–336 (2011).
10. J. K. Gansel *et al*., *Science* **325**, 1513–1515 (2009).
11. A. H. Dorrah, F. Capasso, *Science* **376**, eabi6860 (2022).
12. C. Wu *et al*., *Nat. Commun.* **5**, 3892 (2014).
13. J. Ni *et al*., *ACS Nano* **15**, 2893–2900 (2021).
14. K. Konishi *et al*., *Phys. Rev. Lett.* **106**, 057402 (2011).
15. S. S. Kruk *et al*., *ACS Photonics* **1**, 1218–1223 (2014).
16. A. A. Maksimov *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **89**, 045316 (2014).
17. S. Iba, S. Koh, K. Ikeda, H. Kawaguchi, *Appl. Phys. Lett.* **98**, 081113 (2011).
18. A. Kodigala *et al*., *Nature* **541**, 196–199 (2017).
19. F. Yesilkoy *et al*., *Nat. Photonics* **13**, 390–396 (2019).
20. C. Huang *et al*., *Science* **367**, 1018–1021 (2020).
21. X. Yin, J. Jin, M. Soljačić, C. Peng, B. Zhen, *Nature* **580**, 467–471 (2020).
22. B. Zhen, C. W. Hsu, L. Lu, A. D. Stone, M. Soljačić, *Phys. Rev. Lett.* **113**, 257401 (2014).
23. W. Liu *et al*., *Phys. Rev. Lett.* **123**, 116104 (2019).
24. H. M. Doeleman, F. Monticone, W. den Hollander, A. Alù, A. F. Koenderink, *Nat. Photonics* **12**, 397–401 (2018).
25. K. Koshelev *et al*., *Science* **367**, 288–292 (2020).
26. M. V. Gorkunov, A. A. Antonov, Y. S. Kivshar, *Phys. Rev. Lett.* **125**, 093903 (2020).
27. A. Overvig, N. Yu, A. Alù, *Phys. Rev. Lett.* **126**, 073001 (2021).
28. J. Dixon, M. Lawrence, D. R. BartonIII, J. Dionne, *Phys. Rev. Lett.* **126**, 123201 (2021).
29. I. C. Seo *et al*., *ACS Nano* **15**, 13781–13793 (2021).
30. J. Tian *et al*., *Adv. Mater.* **34**, e2109157 (2022).
31. Materials and methods are available as supplementary materials.

ACKNOWLEDGMENTS

Funding: This work was supported by the National Key Research and Development Project 2021YFA1400802 (Q.S.); Australian Research Council, grant DP210101292 (Y.K.); International Technology Center Indo-Pacific (ITC IPAC) and Army Research Office (contract No. FA520921P0034) (Y.K.); National Natural Science Foundation of China 12025402 (Q.S.); National Natural Science Foundation of China 62125501 (Q.S.); National Natural Science Foundation of China 11934012 (Q.S.); Shenzhen Fundamental Research Projects JCYJ20210324120402006 (Q.S.); and Fundamental Research Funds for Central Universities (Q.S.). **Author contributions:** Conceptualization: Q.S. and Y.K. Methodology: X.Z., Y.L., J.H., and Q.S. Investigation: X.Z. and Q.S. Visualization: X.Z., Q.S., and Y.K. Funding acquisition: Q.S. and Y.K. Project administration: Q.S. Supervision: Q.S. and Y.K. Writing – original draft: Q.S., Y.K., and X.Z. Writing – review and editing: Q.S., Y.K., and J.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq7870
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 and S2
References (*32*–*38*)

Submitted 29 April 2022; accepted 21 July 2022
10.1126/science.abq7870

SOLID-STATE PHYSICS

Spin-selective tunneling from nanowires of the candidate topological Kondo insulator SmB_6

Anuva Aishwarya[1], Zhuozhen Cai[1], Arjun Raghavan[1], Marisa Romanelli[1], Xiaoyu Wang[2], Xu Li[3], G. D. Gu[4], Mark Hirsbrunner[1], Taylor Hughes[1], Fei Liu[3]*, Lin Jiao[1,2]*, Vidya Madhavan[1]*

Incorporating relativistic physics into quantum tunneling can lead to exotic behavior such as perfect transmission through Klein tunneling. Here, we probed the tunneling properties of spin-momentum-locked relativistic fermions by designing and implementing a tunneling geometry that uses nanowires of the topological Kondo insulator candidate samarium hexaboride. The nanowires are attached to the end of scanning tunneling microscope tips and used to image the bicollinear stripe spin order in the antiferromagnet $Fe_{1.03}Te$ with a Neel temperature of about 50 kelvin. The antiferromagnetic stripes become invisible above 10 kelvin concomitant with the suppression of the topological surface states in the tip. We further demonstrate that the direction of spin polarization is tied to the tunneling direction. Our technique establishes samarium hexaboride nanowires as ideal conduits for spin-polarized currents.

A direct consequence of the nontrivial topology of Z_2 topological insulators (TIs) is the existence of an odd number of spin-momentum-locked boundary modes with Dirac dispersion (*1*). When the Fermi energy is tuned inside the bulk band gap, transport is dominated by the Dirac surface states, and marked topological phenomena can be observed. In this scenario, spin-momentum locking coupled with Dirac dispersion dictates that the net current traveling in a given direction must be spin polarized; currents traveling in opposite directions carry opposite spin. The high spin-charge conversion efficiency, and the topological protection of the boundary states, make TIs particularly interesting for applications in "topological spintronics," as well as for quantum technologies such as controlling qubits (*2*–*4*). To date, many exotic phenomena have been realized in magnetically doped TIs or TI heterostructures, such as the quantum anomalous Hall effect (*5*), the topological magnetoelectric effect (*6*), and the large spin–orbit torque effect (*7*). TI boundary modes are also ideal platforms with which to explore the unique tunneling properties of relativistic fermions.

In practice, a substantial bottleneck to using TIs both for fundamental studies and applications is the fact that many semiconductor-based TIs suffer from intrinsic defects that cause self-doping, which moves the Fermi energy into the bulk bands. There has therefore been a tremendous impetus to study interaction-driven topological phases in which the Fermi level is naturally pinned to the topological band gap. To this end, topological Kondo insulators (TKIs) have been proposed as ideal topological systems in which surface Dirac fermions naturally dominate the density of states near Fermi level (*8*). In TKIs, the lattice Kondo effect leads to the emergence of a low-energy, correlation-driven heavy band, which hybridizes with another light band crossing the Fermi energy, leading to a Kondo hybridization gap (*9*). A concomitant band inversion results in the formation of topological surface states (TSS) within this gap. The material samarium hexaboride (SmB_6) (Fig. 1A) has been proposed as a TKI and exhibits the following phenomenology (*9*). At high temperatures, SmB_6 is a Kondo metal. Below the bulk Kondo coherence temperature (T_K) of ~50 K, the hybridization between the emergent heavy f band and the light d bands opens a Kondo gap (Fig. 1B, schematic), which becomes better defined with decreasing temperature (*10*–*14*). Scanning tunneling microscopy (STM) studies have shown that, below 10 to

[1]Department of Physics and Materials Research Laboratory, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. [2]National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, USA. [3]State Key Laboratory of Optoelectronic Materials and Technologies, Guangdong Province Key Laboratory of Display Material and Technology, and School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510275, China. [4]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA.
***Corresponding author. Email: liufei@mail.sysu.edu.cn (F.L.); lin.jiao@magnet.fsu.edu (L.J.); vm1@illinois.edu (V.M.)**

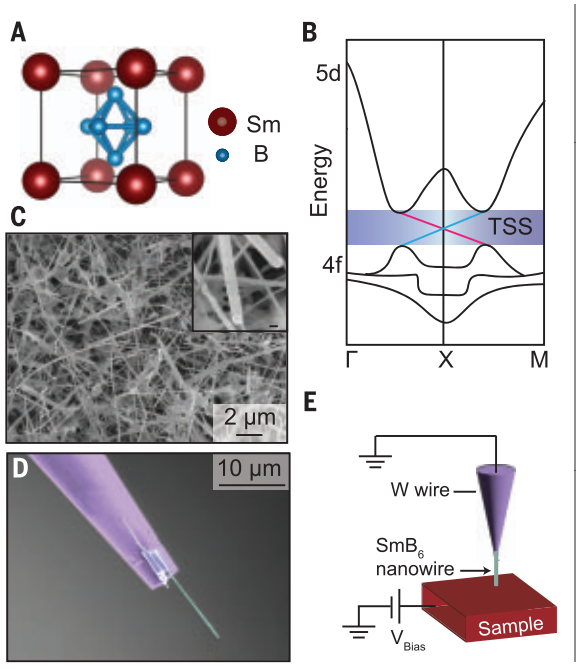


Fig. 1. Morphology, crystal structure, and electronic structure of the SmB_6 nanowire tip. (**A**) Crystal structure of SmB_6 showing the cubic lattice of Sm atoms with octahedrons of boron at the center. (**B**) Schematic of the band structure of SmB_6 showing the hybridization-induced gap between the 5d and 4f bands. The blue shaded region represents the Kondo gap, and the cyan and pink lines indicate the TSS. The colors represent opposite spin orientations. (**C**) High-resolution SEM image of the nanowires on a silicon substrate. Inset: magnified image of a nanowire. Scale bar, 200 nm. (**D**) High-resolution false-colored SEM image of the nanowire tip showing the nanowire bonded to a trimmed W tip. (**E**) Schematic of the STM circuit showing the nanowire tip and the sample.

15 K, "heavy" surface states develop within the Kondo gap (*15*, *16*). The observation of the surface states at far lower temperatures (~10 K) than T_K suggests that the surface Kondo temperature may be lower than T_K (*15*, *17*). The topological nature of the surface states is supported by the observation of a series of phenomena (*18*), e.g., nonlocal transport (*19*, *20*), enhanced spin pumping (*21*), the Edelstein effect (*22*, *23*), chiral edge modes (*24*), spin-momentum locked spin textures for the surface states (*25*), and the Klein paradox (*26*). Despite these studies, the topological origin of the surface states remains controversial (*27*). If the surface states were indeed topological, then the presence of isolated TSS within a bulk band gap that straddles the Fermi energy would make SmB_6 an excellent system for harnessing the properties of Dirac electrons.

In this work, we used a unique tunneling geometry to reveal a distinctive tunneling process that we call helical tunneling. The term "helical" refers to a direction-dependent

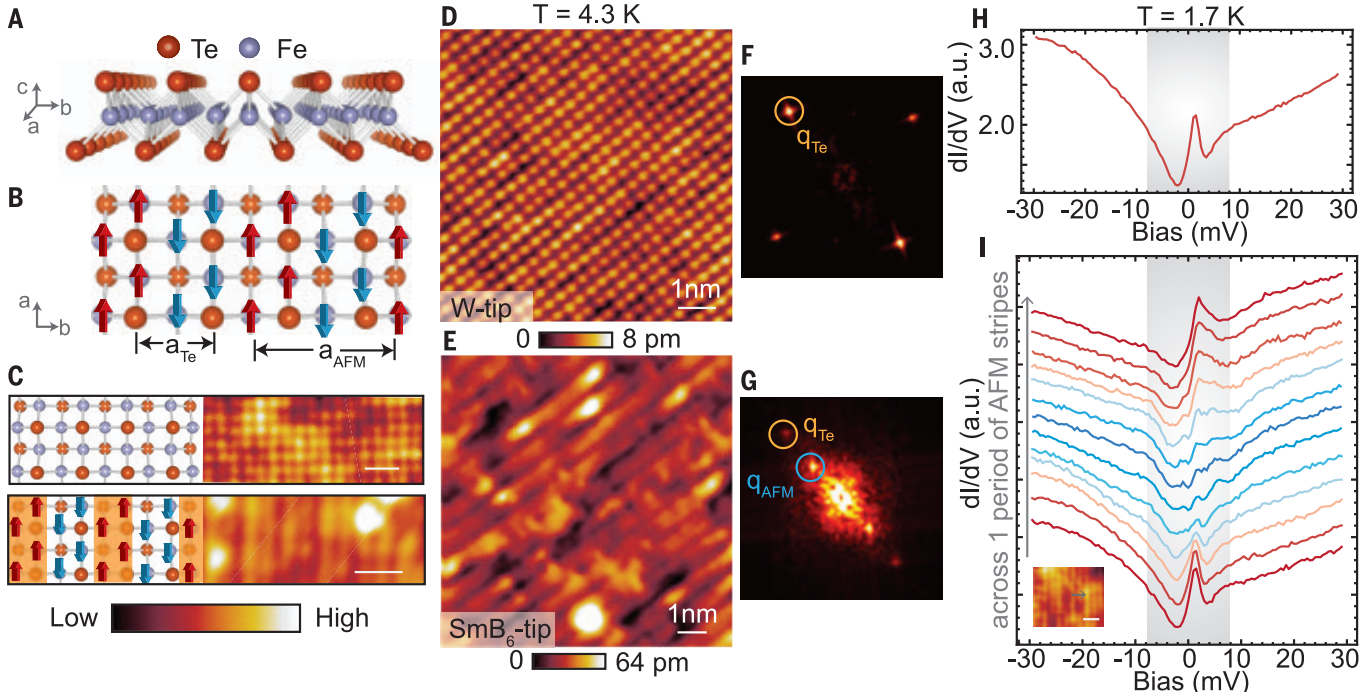


Fig. 2. Atomic-resolution imaging of antiferromagnetic stripes in $Fe_{1+x}Te$ with SmB_6 nanowire tip. (**A**) Crystal structure showing one layer of $Fe_{1+x}Te$ made of an iron (Fe) layer sandwiched between two Te layers. (**B**) Schematic of the *ab* plane of the lattice with the overlay of the spin structure showing the in-plane antiferromagnetic order. The periodicity of the antiferromagnetic order and the lattice are labeled as a_{AFM} and a_{Te}, respectively. (**C**) Schematic of the Te lattice is depicted by the orange spheres (top) and AFM order (bottom). Topography obtained with a W tip (top) and a spin-polarized Cr tip (bottom). The schematic and topography are not the same size. Scale bar in each topography, 1 nm. (**D**) STM topography obtained with a spin-averaging W tip (V_{Bias} = 20 mV, I = 60 pA, T = 4.3 K). (**E**) STM topography obtained with the SmB_6 tip at V_{Bias} = 20 mV, I = 60 pA, T = 4.3 K showing well-defined spin contrast. (**F** and **G**) FFT of the topographies in (D) and (E), respectively, showing only the Bragg peaks from the Te lattice (q_{Te}) in (F) and the presence of a strong signal from the antiferromagnetic ordering (q_{AFM}) along with the q_{Te} in (G). (**H**) Average dI/dV spectroscopy obtained on $Fe_{1+x}Te$ at T = 1.7 K (V_{Bias} = 50 mV, I = 120 pA). In (H) and (I), the highlighted gray region shows the feature in the DOS associated with the topological surface state. (**I**) dI/dV line cut across one period of the AFM stripe (red to blue to red) showing the modulation of the TSS peak as a function of position. Inset shows the topography and the gray arrow shows the line where the spectra were obtained. Scale bar, 1 nm.

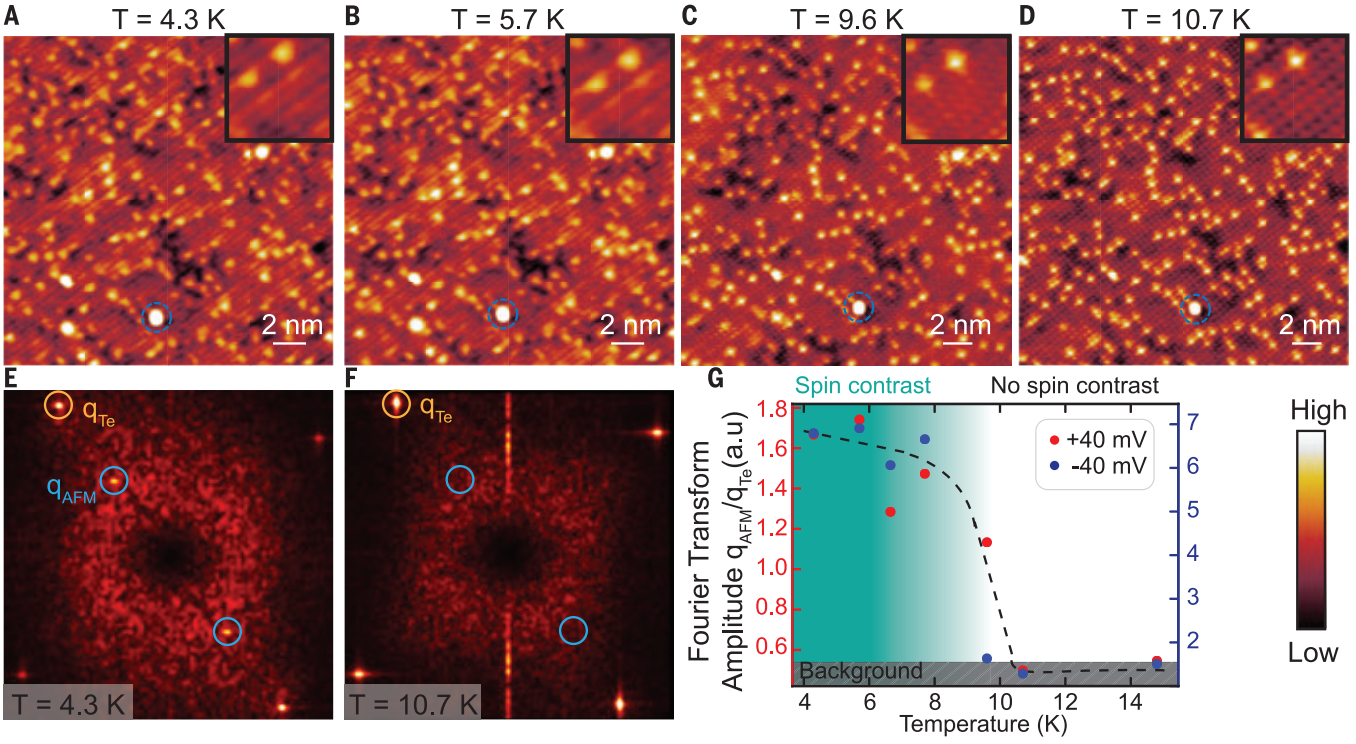


Fig. 3. Temperature dependence showing diminishing spin contrast. (**A** to **D**) STM topography on $Fe_{1+x}Te$ at the labeled temperatures obtained in the same field of view (V_{Bias} = 40 mV, I = 40 pA). Blue circles denote the same defect in all frames. (**E** and **F**) Comparison of FFTs of (A) and (D) to illustrate the disappearance of the q_{AFM} at 10.7 K, far below the antiferromagnetic T_C. A Gaussian background has been subtracted from the center to improve visualization of the q vectors. (**G**) Plot of the ratio of the intensity of the signal from q_{AFM} to the Bragg peak (q_{Te}) as a function of temperature. Black dashed line is a guide to the eye. The spin contrast signal from the AFM stripes is highly suppressed at ~10 K, similar to the behavior of the topological surface state.

spin polarization akin to the spin-momentum locking of helical edge states of a quantum spin Hall insulator. We show that the tunneling electrons from SmB_6 nanowires are spin polarized, with the direction of spin polarization being tied to the direction of tunneling. To achieve this, we used focused ion beam–based nanofabrication techniques to harvest and attach SmB_6 nanowires to the ends of etched tungsten wires typically used as STM tips (see fig. S1) (*28*). The nanowires used in this work were grown on silicon (Si) substrates. They have diameters ranging from 60 to 100 nm and average lengths of ~50 μm (Fig. 1C, inset, and D). The growth procedure is described in detail elsewhere (*29*). The nanowires grow along the [001] direction and terminate laterally at high-symmetry facets such as (001). The larger-diameter nanowires that we used typically contain multiple large facets in which the electronic structure is expected to be similar to the surfaces of bulk SmB_6. Given that the penetration depth of TSS in SmB_6 is <30 nm, our SmB_6 nanowires are sufficiently thick to host TSS (*21*, *26*).

Mounting SmB_6 nanowires on tungsten tips is a two-step process that uses the omniprobe inside the focused ion beam + scanning electron microscopy (SEM) dual-beam system (*28*). The nanowire tips thus prepared are transferred to the STM (Fig. 1E shows a schematic of the tunneling setup). For benchmarking, we first characterized the tips on a Cu(111) single-crystal sample (see figs. S2 and S3) (*28*). Using the facts that the STM dI/dV spectrum is a convolution of tip and sample density of states and that Cu(111) has a flat density of states near the Fermi energy, we could isolate the density of states of the nanowire tip. The data verify that the nanowire tip hosts surface states similar to bulk SmB_6.

On the basis of these observations, we expect the surface states to play a key role in the tunneling characteristics of the nanowire tips. In particular, the spin-momentum locking of topological surface states may result in spin-polarized currents. To determine the spin polarization of the tunnel current, we studied iron telluride ($Fe_{1+x}Te$), which is a bicollinear antiferromagnet (AFM) for x < 0.11 (Fig. 2A). $Fe_{1+x}Te$ has been extensively studied by spin-polarized STM, and it is well established that the atomic-scale AFM order can be observed with a traditional magnetic tip, where it presents itself as stripes with double the tellurium (Te) periodicity (*30*–*32*) (Fig. 2, B and C). Because of the matrix element effects, tunneling is favored between spin parallel states and suppressed between spin antiparallel states (*33*).

For our STM studies, $Fe_{1+x}Te$ (nominal x = 0.03) single crystals were cleaved at ~90 K before inserting them into the STM head. Shown in Fig. 2H is an average dI/dV spectrum obtained on $Fe_{1+x}Te$ at 1.7 K. With SmB_6 on the tip, we observed a peak on the opposite side of E_F compared with spectra on bulk SmB_6 (see comparison with bulk spectrum in fig. S4) (*28*). This demonstrates the existence of the surface states on the tip. We then used this tip to obtain topographic images of $Fe_{1+x}Te$. In contrast to topography using a tungsten (W) tip (Fig. 2D), which shows the expected square Te lattice of the top layer with a lattice constant of 3.8 Å, the images with the SmB_6 nanowire tips show stripes (Fig. 2E). The double periodicity of the observed pattern is the same as the spin modulation of the striped AFM order (Fig. 2G), and the image is consistent with the spin-dependent contrast seen in several previous STM studies using magnetic tips (*30*–*32*). This phenomenology was confirmed using multiple nanowire tips (fig. S5) (*28*). Because it is well established that spin-dependent spatial contrast in STM images arises from the tunneling of spin-polarized electrons, the observed spin contrast directly indicates that

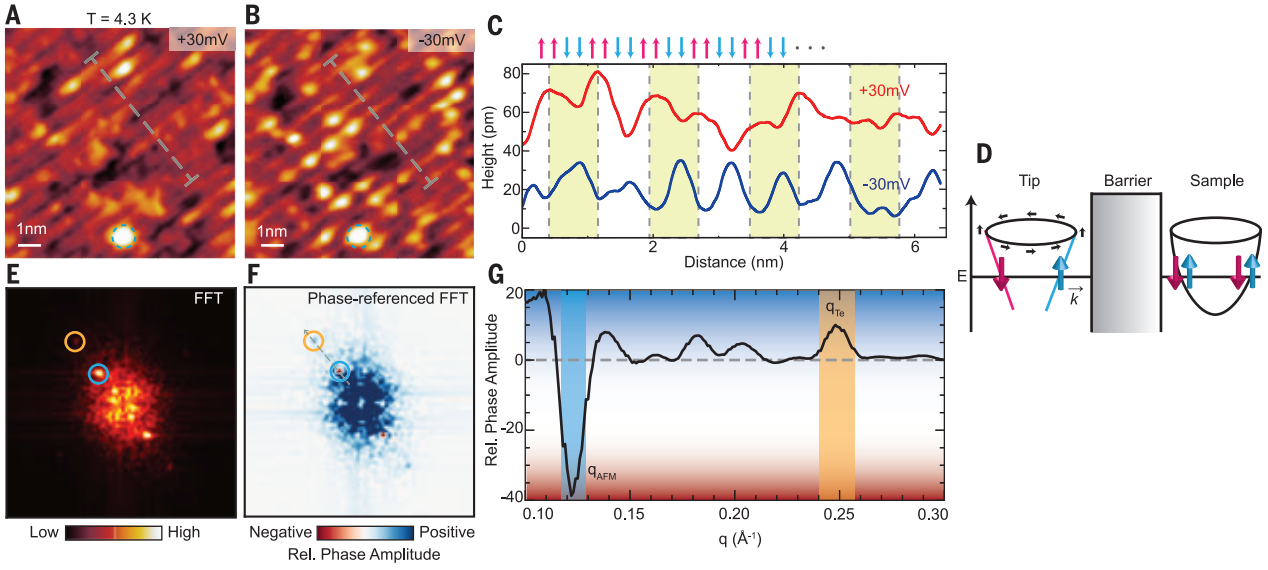


Fig. 4. Helical tunneling with a SmB_6 nanowire tip: Contrast reversal at opposite bias voltages. (**A** and **B**) STM topographies of $Fe_{1+x}Te$ obtained in the same area at +30 mV (−30 mV) and using an SmB_6 tip at T = 4.3 K. (**C**) Height profile of the topography along gray dashed lines in (A) (red curve) and (B) (blue curve). The red curve is shifted vertically for clarity. The dotted lines correspond to the peaks of the red curve and dips of the blue curve, indicating a π-phase shift of the stripes between the positive and negative bias scan. (**D**) Illustration of the tunneling geometry used for the quantum tunneling simulation from the TSS to a normal metal. Without loss of generality, we have assumed the tip to be a 2D metal with just one branch of spin-momentum-locked states crossing the Fermi energy and the sample to be a normal metal with spin-degenerate bands (*28*). In reality, the sample has AFM stripes arising from the Fe lattice, which is visualized by the spin-polarized currents from the tip. (**E**) Absolute magnitude of the FFT of the topography shown in (B) showing the peaks associated with q_{Te} (yellow) and q_{AFM} (cyan). (**F**) Phase-referenced FFT between the topographies shown in (A) and (B), with the same peaks highlighted in cyan and yellow. The relative phase amplitude is negative at q_{AFM}, indicating a π-phase shift of the stripes originating from helical tunneling. (**G**) Line cut across the phase-referenced FFT [indicated by gray dashed line in (F)] illustrating the sign change between q_{Te} and q_{AFM}.

the SmB_6 nanowire tips act as a source of spin-polarized currents.

As a first check of the possible involvement of the surface states in generating the spin-polarized currents, we measured the temperature dependence of the observed spin contrast. Previous transport and spectroscopic experiments on bulk SmB_6 have revealed the attenuation of the Dirac surface states in SmB_6 with increasing temperature and a transition from surface-dominated to bulk-dominated regime around 10 to 14 K (*15*, *16*, *34*, *35*). Figure 3, A to D, shows the temperature-dependent topographies of $Fe_{1+x}Te$ in the same field of view. The AFM order can be seen clearly at lower temperatures. However, at ~10 K, the striped signature of the AFM order disappears, leaving behind only the atomic corrugation of the Te lattice. This is also clearly seen in the fast Fourier transform (FFT) (Fig. 3, E and F). We obtained similar results with two other nanowire tips (fig. S6) (*28*). Upon cooling down, the AFM order reappeared (fig. S7) (*28*). Given that the Neel temperature (T_N) for $Fe_{1.03}Te$ is ~50 to 60 K, the disappearance of magnetic contrast around ~10 K must be caused by a change in the nanowire tip. In fact, previous STM images obtained with magnetic tips continued to show AFM order all the way to 45 K (*30*). In addition, we investigated the same samples with canonical spin-polarized Cr tips, which continued to show clear AFM stripes at 10 K (see fig. S8) (*28*).

Although the temperature dependence suggests a causal relationship between the presence of the surface states and the spin-polarized contrast (fig. S9) (*28*), there may be other explanations for the temperature dependence. For example, metallic Sm is AFM in nature and has a T_N of ~14 K (*36*), so a Sm cluster on the tip might create spin-polarized currents that mimic this temperature dependence. To distinguish between these scenarios, we explored how spin-momentum-locked tunneling of Dirac electrons is different from typical magnetic tips. We calculate the spin-resolved transport coefficients for tunneling between a normal two-dimensional (2D) metal and a 2D band of spin-momentum-locked, singly degenerate linearly dispersing states (Fig. 4D, schematic). A Zeeman term was used to open a gap in the inner band of 2D Rashba states, which mimics tunneling from a 2D Dirac band. The transmission coefficient with an infinitesimal bias voltage is equivalent to the differential conductance, and as is usual in tunneling, the sign of the bias voltage dictates the direction of travel of the tunnel current. The net spin polarization of the tunneling electrons going from left to right ($I_{L\text{-}R}$) and right to left ($I_{R\text{-}L}$) is plotted in fig. S10B (*28*). The calculations show two important features: (i) The tunnel current is indeed spin polarized and (ii) the spin polarization reverses when the direction of tunneling is reversed (helical tunneling). Such a reversal of spin polarization is neither expected nor observed for conventional magnetic tips (*37*), suggesting an unambiguous way to test the involvement of the surface states in generating spin-polarized currents by checking whether the currents are oppositely polarized at opposite bias.

From the discussion above, a "smoking gun" proof of the proposed helical tunneling would be the observation of contrast reversal of the observed AFM stripes at opposite tip bias voltages. Shown in Fig. 4, A and B, and in fig. S11 (*28*) are images in the same area obtained at positive and negative biases with a nanowire tip. The line cuts of the two images (Fig. 4C) show a reversal of the AFM contrast, indicating that the opposite spin channel is selected by the tunneling electrons upon switching the sign of the bias voltage. To better establish that this contrast reversal is pervasive over the whole image, we performed phase-referenced FFT (*38*) (see fig. S12 for the analysis method) (*28*). As shown in Fig. 4, F and G, we observed a negative value for

the relative phase amplitude at the antiferromagnetic ordering vector (q_{AFM}), establishing a π-phase shift of the stripes at opposite biases. This contrast reversal was robust against thermal cycling (fig. S13) (*28*). Moreover, as expected, such contrast reversal was not observed for spin-polarized Cr tips [fig. S14 (*28*) and see (*38*)].

Combined with the temperature dependence, the contrast reversal at opposite bias voltages suggests that the topological surface states participate in generating the spin-polarized tunneling currents observed in our experiment. Although there are different facets on the nanowires, we expect the junction to be formed at the bottom of one of the side crystal planes (see schematic in fig. S15) (*28*). This is important because an equal average over all facets would lead to zero net spin polarization. Indeed, the macroscopic W wire, which hosts the nanowire, is slid into the tip holder and tightened by a set screw, which always causes a tilt, and this naturally results in the tunnel junction being formed at one of the side facets (see schematic in fig. S15) (*28*).

Further support for the involvement of the surface states was obtained by directly measuring how the low-energy features associated with the SmB_6 surface states change across the AFM stripes. Shown in Fig. 2I is a spectroscopic line cut across one period of the AFM stripe. We found that the amplitude of the low-energy features associated with the surface state changed periodically as the tip was moved across an AFM stripe. This change in intensity is a result of changes in the tunneling probability caused by the different relative orientations of the tip-surface state electron spin and the sample spin. This scenario of TSS-mediated helical tunneling will survive even with the weak surface magnetism that has been observed in SmB_6 at millikelvin temperatures (*24*). It has been suggested that the TSS may mediate a ferromagnetic RKKY interaction among magnetic impurities on the surface, giving rise to this surface magnetism (*39*). However, because the Dirac surface states are generated by the nontrivial topology of the bulk, weak surface magnetism would simply open a small (<1 meV) gap at the Dirac point, leaving the remainder of the helical tunneling physics outside of this energy window unchanged. This would not be at odds with any of our observations.

In summary, the reduced dimensionality and geometry of the nanowire tips allowed us to study the tunneling properties of a candidate TKI. Our observations and calculations show that the spin polarization of the tunnel current is insensitive to the position of the Fermi energy relative to the Dirac point and only depends on the direction of tunneling. These unique properties of spin-momentum-locked Dirac fermions may be used directly for manipulating spin currents in the field of spintronics. In addition, magnetic contrast at surfaces or local defects can be detected with atomic resolution without using external magnetic fields, as would be necessary in a traditional spin-polarized STM. Equally important, our technique of mounting nanowires on the STM tip opens up avenues for using topological nanowire tips for devices and heterostructures (*40*, *41*). As an example, a proximitized topological nanowire in a vertical magnetic field would host Majorana fermions at the end, effectively making a Majorana tip. This may then be used for directly studying interactions as a function of distance between two Majorana modes (one on the nanowire tip and the other on the sample), which has proven difficult to do so far. Finally, scanning probe microscopy using unconventional tips is a burgeoning field in the study of quantum materials for nanoscale magnetic imaging (*42*, *43*). Our work thus opens many avenues for using various kinds of functionalized nanowires as STM tips (superconducting, topological, etc.) for atomic-resolution imaging to provide nanoscale information on different materials properties such as magnetism, superconductivity, and other emergent excitations.

REFERENCES AND NOTES

1. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
2. K. L. Wang, M. Lang, X. Kou, "Spintronics of topological insulators," in *Handbook of Spintronics*, Y. Xu, D. Awschalom, J. Nitta, Eds. (Springer, 2016).
3. L. Šmejkal, Y. Mokrousov, B. Yan, A. H. MacDonald, *Nat. Phys.* **14**, 242–251 (2018).
4. Y. Tokura, K. Yasuda, A. Tsukazaki, *Nat. Rev. Phys.* **1**, 126–143 (2019).
5. C.-Z. Chang *et al*., *Science* **340**, 167–170 (2013).
6. K. N. Okada *et al*., *Nat. Commun.* **7**, 12245 (2016).
7. Y. Fan *et al*., *Nat. Mater.* **13**, 699–704 (2014).
8. M. Dzero, K. Sun, V. Galitski, P. Coleman, *Phys. Rev. Lett.* **104**, 106408 (2010).
9. M. Dzero, J. Xia, V. Galitski, P. Coleman, *Annu. Rev. Condens. Matter Phys.* **7**, 249–280 (2016).
10. X. Zhang *et al*., *Phys. Rev. X* **3**, 011011 (2013).
11. S. Rößler *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4798–4802 (2014).
12. M. Neupane *et al*., *Nat. Commun.* **4**, 2991 (2013).
13. W. Ruan *et al*., *Phys. Rev. Lett.* **112**, 136401 (2014).
14. W. K. Park *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6599–6604 (2016).
15. L. Jiao *et al*., *Nat. Commun.* **7**, 13762 (2016).
16. H. Pirie *et al*., *Nat. Phys.* **16**, 52–56 (2020).
17. V. Alexandrov, P. Coleman, O. Erten, *Phys. Rev. Lett.* **114**, 177202 (2015).
18. L. Li, K. Sun, C. Kurdak, J. W. Allen, *Nat. Rev. Phys.* **2**, 463–479 (2020).
19. S. Wolgast *et al*., *Phys. Rev. B* **88**, 180405 (2013).
20. D. J. Kim *et al*., *Sci. Rep.* **3**, 3150 (2013).
21. T. Liu *et al*., *Phys. Rev. Lett.* **120**, 207206 (2018).
22. A. R. Mellnik *et al*., *Nature* **511**, 449–451 (2014).
23. Q. Song *et al*., *Nat. Commun.* **7**, 13485 (2016).
24. Y. Nakajima, P. Syers, X. Wang, R. Wang, J. Paglione, *Nat. Phys.* **12**, 213–217 (2016).
25. N. Xu *et al*., *Nat. Commun.* **5**, 4566 (2014).
26. S. Lee *et al*., *Nature* **570**, 344–348 (2019).
27. P. Hlawenka *et al*., *Nat. Commun.* **9**, 517 (2018).
28. Materials and methods and supporting data are available as supplementary materials.
29. H. Gan *et al*., *Cryst. Growth Des.* **19**, 845–853 (2019).
30. M. Enayat *et al*., *Science* **345**, 653–656 (2014).
31. T. Hänke *et al*., *Nat. Commun.* **8**, 13939 (2017).
32. S. Manna *et al*., *Nat. Commun.* **8**, 14074 (2017).
33. R. Wiesendanger, *Phys. Rev. Lett.* **81**, 1495 (2009).
34. J. D. Denlinger *et al*., Temperature dependence of linked gap and surface state evolution in the mixed valent topological insulator SmB_6. https://arxiv.org/abs/1312.6637 (2013).
35. F. Chen *et al*., *Phys. Rev. B* **91**, 205133 (2015).
36. M. Mito *et al*., *Phys. Rev. B* **104**, 054431 (2021).
37. M. Enayat, "Construction and operation of a milli-kelvin spectroscopic imaging STM for the study of correlated electron materials, PhD thesis, Swiss Federal Institute of Technology Lausanne (2014); https://www.fkf.mpg.de/5191581/dok92-Mostafa_Enayat_Thesis_2014.pdf.
38. S. Chi *et al*., Determination of the superconducting order parameter from defect bound state quasiparticle interference. arXiv:1710.09089 [cond-mat.supr-con] (2017).
39. Q. Liu, C.-X. Liu, C. Xu, X.-L. Qi, S.-C. Zhang, *Phys. Rev. Lett.* **102**, 156603 (2009).
40. P. Liu, J. R. Williams, J. J. Cha, *Nat. Rev. Mater.* **4**, 479–496 (2019).
41. R. Lutchyn *et al*., *Nat. Rev. Mater.* **3**, 52–68 (2018).
42. F. Casola, T. van der Sar, A. Yacoby, *Nat. Rev. Mater.* **3**, 17088 (2018).
43. D. Vasyukov *et al*., *Nat. Nanotechnol.* **8**, 639–644 (2013).
44. Raw files and code for: A. Aishwarya *et al*,, Spin-selective tunneling from nanowires of the candidate topological Kondo insulator SmB6, Illinois Data Bank (2022); https://doi.org/10.13012/B2IDB-9971603_V1.

ACKNOWLEDGMENTS

We thank H. Zhou, Senior Research Scientist, Materials Research Laboratory, University of Illinois, for help in using the focused ion beam + SEM dual-beam system and S. Wirth and P. Xiong for helpful discussions. **Funding:** The STM work at the University of Illinois was supported by the National Science Foundation through grant DMR-2003784. M.H and T.L.H thank the Army Research Office (ARO) under Multidisciplinary University Research Initiatives (MURI) grant W911NF2020166 for support. V.M acknowledges partial support from the Gordon and Betty Moore Foundation (EPiQS grant 9465) and the Materials Research Laboratory Central Research Facilities at the University of Illinois for use of the cleanroom. V.M is a CIFAR Fellow in the Quantum Materials Program and acknowledges CIFAR for support. A.A. was partially funded by PPG-Materials Research Laboratory Graduate assistantship 2020-2021. X.L and F.L. are very thankful for the support of the National Science Foundation of China (grant 51872337) and the National Project for the Development of Key Scientific Apparatus of China (grant 2013YQ12034506). L.J. and X.W. acknowledges financial support from National MagLab, which is funded by the National Science Foundation (grant DMR-1644779) and the State of Florida. **Author contributions:** A.A., L.J., and V.M. conceived the experiments. A.A., A.R., and L.J. obtained the STM data and characterized the tips and samples. L. J., M.R., and Z.C. fabricated the nanowire tips. G.D.G. provided the single crystals of $Fe_{1.03}Te$. X.L. and F.L. grew the nanowires of SmB_6. X.W., M.H., and T.H. performed the theoretical modeling. A.A., L.J., and V.M. performed the data analysis and wrote the manuscript with help from all the authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data have been uploaded to the Illinois Data Bank (*44*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj8765
Materials and Methods
Supplementary Text
Figs. S1 to S16
References (*45*–*47*)

Submitted 7 June 2021; resubmitted 8 January 2022
Accepted 5 August 2022
10.1126/science.abj8765

HYDROFORMYLATION

Hydroformylation catalyzed by unmodified cobalt carbonyl under mild conditions

Baoxin Zhang[1]*, Christoph Kubis[1], Robert Franke[2,3]*

Hydroformylation with unmodified cobalt carbonyl catalyst plays a crucial role in industrial production of surfactants and plasticizers. However, syngas pressures of 100 to 400 bar with reaction temperatures of 100° to 250°C are typically applied. We report here that unmodified cobalt carbonyl is a stable hydroformylation catalyst at 140°C under 30 bar of syngas. The activity was comparable to that of recently reported bisphosphine-coordinated cobalt(II) catalysts, which we could not reproduce under the reported conditions. Kinetic and in situ infrared spectroscopic studies confirmed the stability of the unmodified cobalt tetracarbonyl hydride [$HCo(CO)_4$]. Branched internal olefins were converted to aldehydes with high regioselectivity under low syngas pressures without phosphorus ligands. Bisphosphines had a small promotional effect on the catalyst at phosphorus-to-cobalt loading ratios below 0.6.

The preferred route to aldehydes in industrial organic chemistry is hydroformylation, the catalytic addition of synthesis gas (a mixture of H_2 and CO) to olefins (*1*). The resulting aldehydes are starting materials for numerous important bulk and fine chemicals (*2*–*4*). The reaction is homogeneously catalyzed by cobalt and rhodium complexes (*5*). The hydroformylation of short olefins is nowadays carried out almost exclusively using rhodium catalysts modified with phosphorus ligand, which achieve high chemo- and regioselectivities under comparatively mild conditions (*6*–*10*). However, processes based on cobalt catalysts dominate the production of aldehydes from longer-chain, especially highly branched olefins with internal double bonds (*11*).

At reaction temperatures (usually 100° to 250°C), the stability of the active catalyst, $HCo(CO)_4$, is sensitive to the CO partial pressure (*12*, *13*). Thus, the applied syngas (H_2:CO = 1:1) pressure in industry is between 100 and 400 bar. It is believed that under the reaction conditions, $HCo(CO)_4$ decomposes through the loss of one coordinated carbonyl group, which leads to combination with $HCo(CO)_4$ at the free coordination site (*14*). Release of H_2 and coordination by another CO ligand then produces $Co_2(CO)_8$. The cobalt dimer can also catalyze the dimerization of $HCo(CO)_4$ through a radical mechanism (*15*–*18*). If the CO partial pressure is insufficient to stabilize the accumulated $Co_2(CO)_8$, this complex will lose further CO ligands to produce inactive multinuclear Co clusters and even Co metal.

Phosphine ligands can substantially improve the stability of the cobalt complexes (*19*) but nevertheless also decelerate the catalysis (*20*). Higher temperatures, usually 160° to 200°C, are required to activate the dissociation of carbonyl from the hydridocobalt complex, initiating the catalytic cycle. The electron-rich phosphines also increase the hydrogenation tendency of the catalyst. These drawbacks were putatively overcome recently by a cationic Co(II) system modified with bisphosphines, reported by Stanley and coworkers (*21*). This Co complex catalyzed hydroformylation of 1-hexene under comparatively mild conditions (50 bar, 140°C). We sought to repeat the same experiments in our laboratory. However, we could not reproduce the results from Stanley's group. Unexpectedly, we found that unmodified $HCo(CO)_4$, which is commonly considered unstable under 30 to 50 bar of syngas at 140°C, could perform hydroformylation under these mild conditions, without appreciable decomposition during the reaction. The new system has a high tolerance for cobalt precursors. Common cobalt sources of various oxidation states—e.g., $Co_2(CO)_8$, $Co(acac)_2$, etc.—are suitable to generate the active $HCo(CO)_4$. Under the same conditions as those of Stanley's system, the unmodified catalyst showed very similar activity and *n*-selectivity for the same substrate.

Because in industrial practice, the unmodified cobalt catalyst is mostly applied in the hydroformylation of internal olefins with long and branched carbon chains, for which usually harsh conditions (i.e., pressures above 100 bar and temperatures above 150°C) are required owing to the low activity of these substrates, studies reporting catalysis with phosphorus-free cobalt carbonyls under mild conditions are rare. Until now only one patent published in 1956 mentioned that the unmodified cobalt catalyst was still active below 50-bar syngas pressure when the reaction temperatures were between 90° and 150°C (*22*). Nevertheless, a catalyst preformation step is required at higher pressures of 175 to 250 bar. The capacity of $HCo(CO)_4$ to catalyze hydroformylation at low

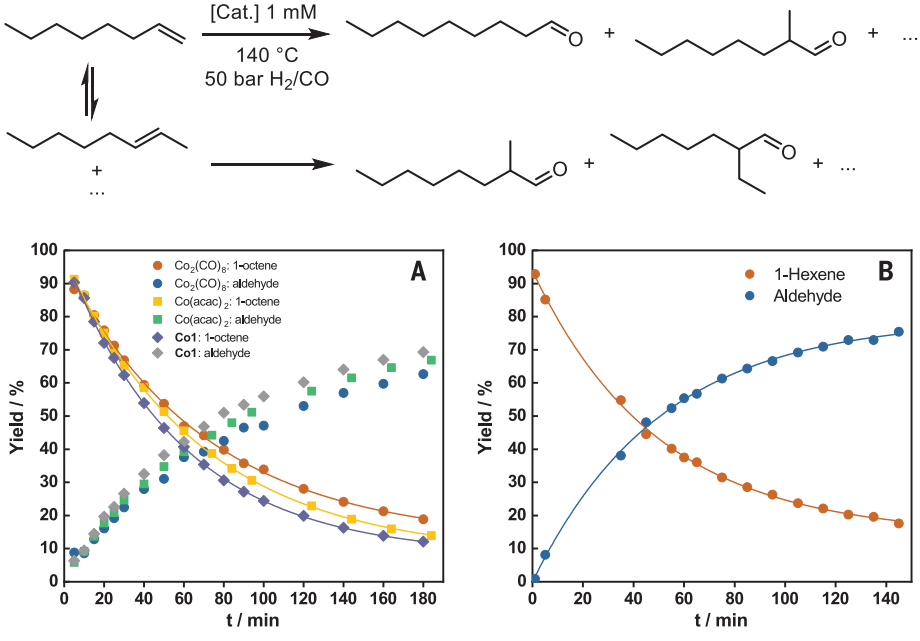


Fig. 1. Conversions of substrate and product formation with different catalyst precursors for hydroformylation. Substrate: 1-octene (**A**), 1-hexene (**B**). Conditions: 1 mM Co (61 ppm), 1 M olefin, 140°C, 50 bar syngas (H_2:CO = 1:1). Catalyst preformation at 160°C, 50 bar syngas for 30 min. For (B), 1-hexene was dissolved in tetraglyme with $Co(acac)_2$ as precursor. Dots: experimental data; lines: exponential fits. The complete experimental data are listed in tables S4 to S7.

[1]Leibniz Institute for Catalysis e.V., Albert-Einstein-Str. 29a, 18059 Rostock, Germany. [2]Evonik Operations GmbH, Paul-Baumann-Str. 1, 45772 Marl, Germany. [3]Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany.
*Corresponding author. Email: baoxin.zhang@catalysis.de (B.Z); robert.franke@evonik.com (R.F.)

pressure was reported in the interim by academic groups. Banerjee *et al.* reported in 1981 that cyclohexene was hydroformylated in the presence of the unmodified cobalt catalyst at 50 bar and 130°C, producing the corresponding aldehyde in 20% yield (*23*). Bourne and coworkers showed in 1992 that using $Co_2(CO)_8$ as precatalyst, propylene could be converted to aldehyde at CO partial pressures in the 10- to 75-bar range with a constant H_2 partial pressure of 25 bar, at 110° to 150°C (*24*). Hydroformylation of 1-pentene (*25*, *26*) and 1-hexene (*20*, *27*, *28*) by $Co_2(CO)_8$ under 30 to 50 bar of syngas was reported with activities that were dependent on solvent, catalyst concentration, temperature, and reaction time. The unmodified cobalt carbonyl was also moderately active in hydroformylation of functionalized alkenes such as vinyl acetate (*29*) and 3,4-dihydro[2H]pyran (*30*) under mild conditions. However, none of these previous studies thoroughly characterized the active catalytic species or the full extent of its stability. We rigorously establish herein the underappreciated stability of the unmodified hydridocobalt catalyst.

A variety of cobalt sources can serve as catalyst precursors, converted in solution to $HCo(CO)_4$ upon reaction with syngas (Fig. 1). We have chosen three precursors that are representative of common cobalt species of various oxidation states and charges: $Co_2(CO)_8$, $Co(acac)_2$ ($acac^-$ = acetylacetonate anion), and $[Co_m(acac)_m(C_4H_8O_2)_n](BF_4)_m$ (**Co1**). The latter was prepared by adding one equimolar HBF_4 into a $Co(acac)_2$ solution in 1,4-dioxane (*21*), which is similar to the cationic cobalt precursor used in the system of Stanley's group. Its structure and preparation procedure are shown in the supplementary materials (section Experimental). Their catalytic performances are summarized in Table 1.

The unmodified cobalt catalyst produces the aldehyde from 1-hexene in a yield of 74.1% within 2 hours (entry 4), 56.5% of which was transformed after 1 hour at a linear-to-branched (*l*/*b*) ratio of 1.5. The activity and *n*-selectivity remained the same independent of the catalyst precursor under conserved conditions. There was no decomposition of the catalyst during the reactions, according to an in situ infrared (IR) spectroscopic reaction monitoring (Fig. 2). These results show a catalytic performance comparable to that of the cationic Co(II) bisphosphine catalyst recently reported by Stanley and co-workers (*21*). Nevertheless, they contradict Stanley's finding that $HCo(CO)_4$ decomposes to cobalt metal under the "medium-pressure conditions" reported there, which are the same conditions applied here. We will address the identification of $HCo(CO)_4$ in our system and its stability under the mild conditions in the following parts of this article. Furthermore, the catalyst is also active for mixtures of internal olefins (entry 16). The *n*-selectivity remained almost unchanged, which we attribute to the strong isomerization activity of the cobalt complex (*31*).

Lower temperature favors linear aldehydes in 1,4-dioxane (entries 8 to 10, Table 1), which is in accordance with the observation of Stanley and co-workers. According to the mechanism proposed by Heck and Breslow (*32*, *33*), the catalytic cycle starts with the carbonyl dissociation from $HCo(CO)_4$, affording $HCo(CO)_3$. From the latter, an alkyl cobalt complex $RCo(CO)_4$ is formed after CO addition, the structure of which is energetically more favorable for the linear product. (*34*) Meanwhile, a decreased CO solubility caused by lower temperatures (*35*) facilitates β-H elimination of the alkyl complex (*36*, *37*). As a result, 1-octene was released from the isoalkyl complex. The better *n*-selectivity is therefore ascribed to both thermodynamic and kinetic factors. Lower CO concentrations not only shift the equilibrium between $HCo(CO)_4$ and $HCo(CO)_3$ in the reaction solution (*13*), but also alter the relative rates of the alkene isomerization and carbon monoxide addition and insertion.

We undertook kinetic studies to confirm the assumption that the same active catalyst, namely $HCo(CO)_4$, accounts for the hydroformylation results under our mild conditions, in spite of the different precursors employed. Concentration profiles of substrate as well as products of the hydroformylation of 1-octene were investigated by reaction sampling and gas chromatography analysis. Three precursor–solvent systems were selected for comparison: (i) $Co_2(CO)_8$ in *n*-heptane; (ii) $Co(acac)_2$ in 1,4-dioxane; and (iii) **Co1** in diglyme. The results are shown in Fig. 1A.

To ensure that the catalyst was fully activated, a 30-min preformation was conducted by treating the precursor with 50-bar syngas at 160°C before alkene injection. (The preformation period was determined through in situ IR spectroscopy, Fig. 2C.) The system with $Co_2(CO)_8$ in *n*-heptane was chosen as the benchmark, because it undoubtedly forms $HCo(CO)_4$. As expected, the rates of olefin consumption and aldehyde production of the three systems differ only slightly, which we attribute to variations of gas solubility in different solvents. The fact that similar yields and *n*-selectivities were obtained at each time point supports our conjecture that all the three precursors lead to the same active species, $HCo(CO)_4$. An observed first-order decay of olefin gives k_{obs} of 0.0140, 0.0146, and 0.0171 min^{-1} with the adjusted coefficient of determination R^2 of 0.99865, 0.99957, and 0.99968, respectively. The initial turnover frequencies (TOFs) determined at the second minute are 11.3 min^{-1} for $Co_2(CO)_8$ in *n*-heptane, 12.8 min^{-1} for $Co(acac)_2$ in 1,4-dioxane, and 14.9 min^{-1} for **Co1** in diglyme. All are in the typical range of TOFs for the unmodified Co(I) catalyst (5 to 20 min^{-1}) (*21*).

Table 1. Hydroformylation of linear olefins by cobalt catalysts. Conditions: 50-bar syngas (H_2:CO = 1:1). The catalyst was preformed at 160°C under 50-bar syngas for 5 min. Then the reaction mixture was cooled to operating temperatures before substrate injection.

Entry	Precursor	Substrate	Solvent	[Co] (mM)	*T* (°C)	*t* (min)	Aldehyde (%)	*n*-Aldehyde (%)
1*	$Co_2(CO)_8$	1-Octene	*n*-Heptane	1	140	180	62.7	54.7
2	$Co_2(CO)_8$	1-Octene	Diglyme	2.6	140	60	63.4	55.1
3	$Co_2(CO)_8$	1-Octene	Toluene	2.6	140	60	61.5	57.4
4*	$Co(acac)_2$	1-Hexene	Tetraglyme	1	140	120	74.1	60.2
5*	$Co(acac)_2$	1-Octene	1,4-Dioxane	1	140	180	66.9	61.0
6	$Co(acac)_2$	1-Octene	1,4-Dioxane	1	140	60	48.1	57.9
7	$Co(acac)_2$	1-Octene	1,4-Dioxane	1.9	140	60	61.1	57.3
8	$Co(acac)_2$	1-Octene	1,4-Dioxane	2.6	140	60	64.7	57.6
9	$Co(acac)_2$	1-Octene	1,4-Dioxane	2.6	120	180	63.7	65.2
10	$Co(acac)_2$	1-Octene	1,4-Dioxane	2.6	110	360	62.7	68.7
11*	**Co1**	1-Octene	Diglyme	1	140	180	69.3	57.4
12	**Co1**	1-Octene	Diglyme	1	140	60	45.3	55.4
13	**Co1**	1-Octene	Diglyme	1.6	140	60	55.8	55.0
14	**Co1**	1-Octene	Diglyme	1.9	140	60	59.4	54.8
15	**Co1**	1-Octene	Diglyme	2.6	140	60	63.5	54.5
16	**Co1**	*n*-Octenes†	Diglyme	2.6	140	240	80.5	51.7
17	None	1-Octene	Diglyme	0	140	60	2.2	-

*The catalyst was preformed at 160°C and 50-bar syngas for 30 min. †3.3% 1-octene, 48.4% *Z*/*E*-2-octene, 29.2% *Z*/*E*-3-octene, 16.4% *Z*/*E*-4-octene, 2.1% skeletal C_8-olefinic isomers, 0.6% *n*-octane.

This notable activity of the unmodified cobalt catalyst under mild conditions prompted us to conduct a direct comparison with the recently published bisphosphine-modified Co(II) catalyst by Stanley and co-workers (*21*). As illustrated in Fig. 1B, under the same conditions, the unmodified Co(I) catalyst afforded very similar activities to those of the phosphine-coordinated Co(II) catalyst. A first-order exponential fit to the decay of 1-hexene provided a k_{obs} of 0.0202 min^{-1} (adjusted coefficient of determination R^2 of 0.9988) and an initial TOF = 15.5 min^{-1} at the second minute. Compared with the aldehyde yields of 71.3 and 51.7% reported after 1 hour with the Co(II) catalyst, the 1-hour-yield of aldehyde was 56.5%, with a *l*/*b* ratio of 1.5 (1.3 for the Co(II) system). A total of 74.1% of olefin was converted to aldehyde with our unmodified catalyst after 2 hours.

An absolutely unexpected result is that the cationic Co(II) catalysts **Co4** and **Co6** (Fig. 3), which we synthesized by the procedure of Stanley and co-workers, have extremely low activity under the published conditions (table S2). **Co4** and **Co6** (preparation procedure and analytic data are in the supplementary materials) are two of the same pre-prepared cationic cobalt catalysts consisting of a Co(II)-center coordinated by a bisphosphine ligand reported by Stanley's group (*21*). Figure S33 illustrates a direct comparison of the hydroformylation performances using **Co4** and the unmodified $Co(acac)_2$ as the catalyst precursor, respectively. For the sake of comparison, the substrate and solvent applied were the same as in Stanley's work. Concerning the temperature effect on preformation rate (see discussion in the following paragraph), we applied a preformation period of 100 min under 50bar of syngas at 160°C to ensure that the catalyst was fully activated. The reaction mixture was then cooled to 140°C before olefin was injected. Only 3.1% of 1-hexene was converted to aldehyde after 20 hours. Attributing the inactivity to the long incubation period does not seem convincing, as this catalyst was reported to be exceptionally stable under the same conditions, according to that publication. The inactivity of this bisphosphine-modified cobalt catalyst is, however, in good accordance with our findings about the ligand effect (table S15 and fig. S34). For substrates other than 1-hexene—for example, 1-octene—the poor activity of this Co(II) cationic precatalyst is shown in comparison with the in situ–formed modified cobalt system with the same ligand (entries 6 to 7, table S15).

Concrete evidence for the formation of $HCo(CO)_4$ from the precursors $Co(acac)_2$ and **Co1** was provided by in situ IR experiments that monitored the catalyst preformation at 160°C. Each precursor was studied in two autoclaves of different volumes, which were

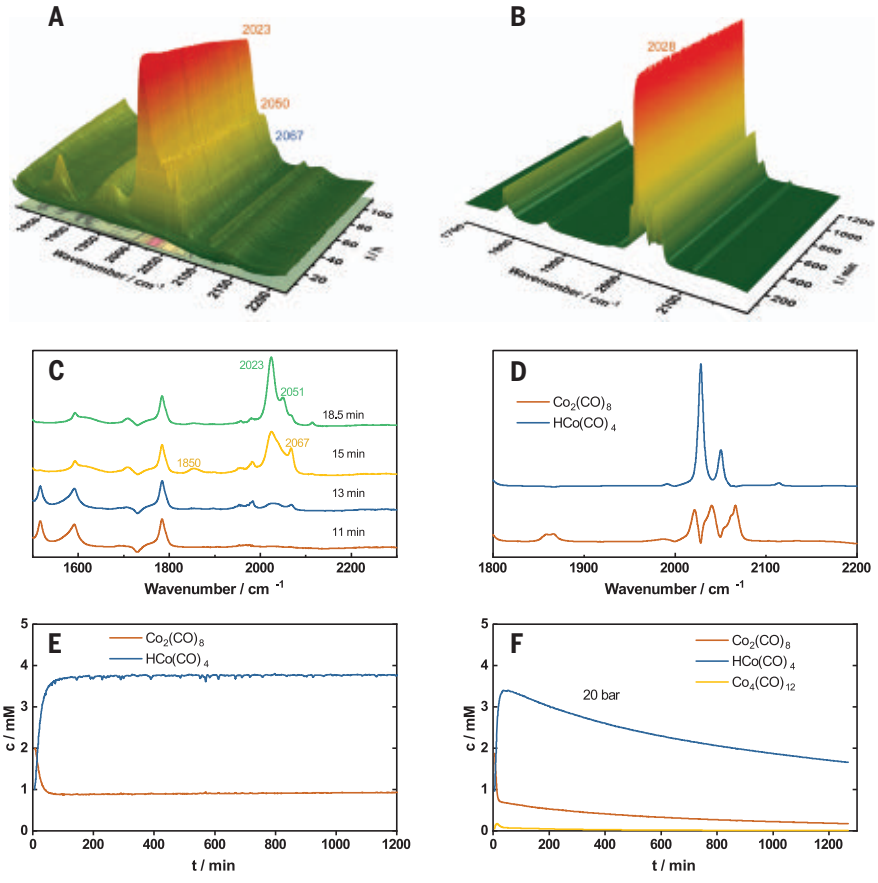


Fig. 2. In situ IR spectroscopic investigation of unmodified cobalt carbonyls. (**A**) Time-resolved IR spectra of preformation and prolonged stability test of $HCo(CO)_4$. Conditions: 5 mM $Co(acac)_2$ in tetraglyme (285 ppm cobalt), 10 mM diphenyl carbonate as internal standard, 140°C, 50 bar syngas (H_2:CO = 1:1). Bands at ν = 2023 and 2050 cm^{-1} are assigned to $HCo(CO)_4$, and ν = 2067 cm^{-1} to $Co_2(CO)_8$. (**B**) Stability test of $HCo(CO)_4$ at a low syngas pressure. Conditions: 2.5 mM $Co_2(CO)_8$ in *n*-dodecane ([Co] = 5 mM), 140°C, 20 bar syngas. (**C**) IR spectra of preformation of $HCo(CO)_4$ in a 200-ml autoclave. Conditions: 5 mM $Co(acac)_2$ in 1,4-dioxane (285 ppm cobalt), 10 mM diphenyl carbonate (ν(CO) = 1784 cm^{-1}) as internal standard, 160°C, 50 bar syngas. Bands at ν = 2023 and 2051 cm^{-1} are assigned to $HCo(CO)_4$, 1850 and 2067 cm^{-1} to $Co_2(CO)_8$, 1516 and 1591 cm^{-1} to $acac^-$. (**D**) Single-component spectra extracted from (B) by Peak Group Analysis (PGA) (*38*, *39*). (**E**) Concentrations of components observed during (B). (**F**) Concentrations of species formed in the stability test of $HCo(CO)_4$ under 20 bar of syngas at 160°C. Conditions: 2.5 mM $Co_2(CO)_8$ in *n*-dodecane ([Co] = 5 mM). Corresponding time-resolved IR spectra and single-component spectra extracted by PGA are in the supplementary materials.

Table 2. Hydroformylation of internal and branched olefins with $Co(acac)_2$ in 1,4-dioxane. Conditions: 1 mM catalyst (61 ppm cobalt), 1 M substrate, 30 bar syngas (H_2:CO = 1:1). Catalyst was preformed at 140°C, 30 bar syngas for 2 hours before olefin injection.

Entry	Substrate	*P* (bar)	*T* (°C)	*t* (hours)	Aldehyde (%)	*n*-Aldehyde (%)	Alcohol (%)
1		30	140	4	15.9	100	0
2		30	140	20	79.6	50	7.5
3		30	140	20	53.7	100	4.0

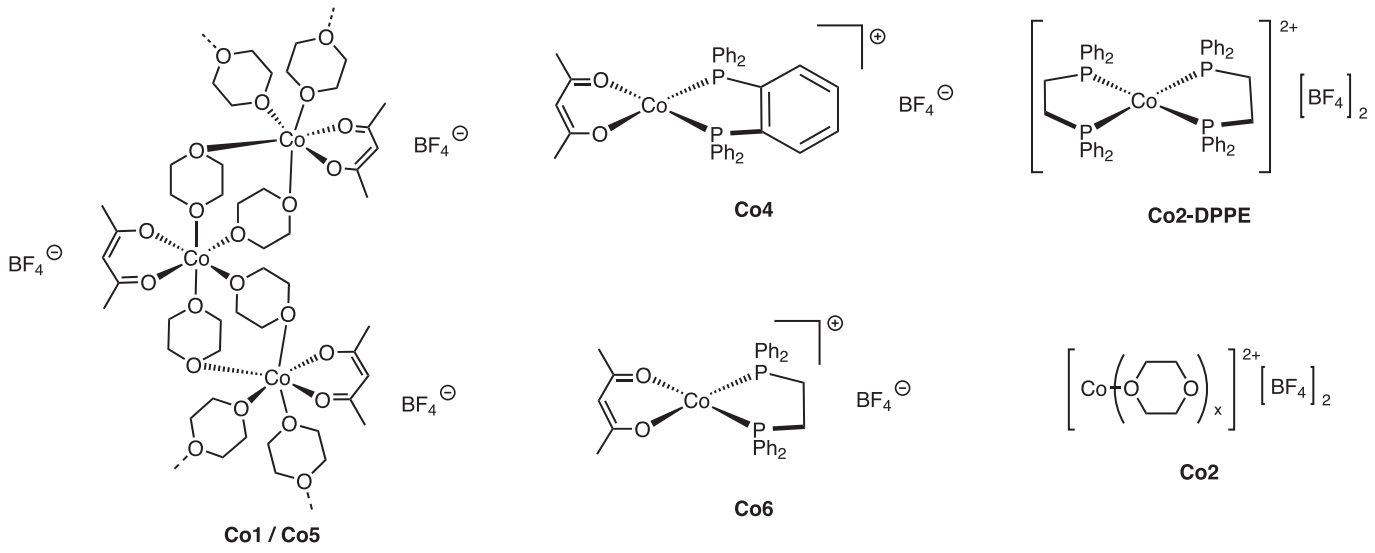


Fig. 3. Structures of the cobalt catalyst precursors in this study.

connected via capillaries to a high-pressure transmission Fourier transform infrared (FTIR) cell. The reaction solution was circulated by a microgear-pump between the reactor and FTIR cell, the latter of which was heated at a temperature of 120°C for protecting the window material during the measurement (for details, see fig. S21). Rate of formation of $HCo(CO)_4$ proved very sensitive to temperature. Despite good thermal insulation, temperature differences between the solution inside the autoclave and in the measuring periphery were unavoidable. Therefore, the real reaction temperature in the large reactor (200 ml with 45 ml of solution) was higher than in the small one (25 ml with 15 ml of solution), given that the volume ratio between reactor and periphery is greater for the former. Consequently, whereas no formation of $HCo(CO)_4$ was traced until 1 hour of reaction time in the small scale autoclave (50-bar syngas, 160°C), in the large reactor, where the aforementioned kinetic studies were carried out, catalyst preformation from both precursors was fully complete within 20 min. Two characteristic carbonyl bands of $HCo(CO)_4$, ν(CO) = 2024 and 2050 cm^{-1}, were observed in the IR spectra. $Co_2(CO)_8$ was detected at the outset (bands at ν(CO) = 1850, 2038, and 2067 cm^{-1}), and then its concentration decayed rapidly until it became undetectable after 6 min (Fig. 2C). By contrast, carbonyl bands of $Co_2(CO)_8$ were always present after the catalyst activation in the small reactor, although indicative of relatively low concentrations in comparison to $HCo(CO)_4$. The catalyst formation process, as well as relative concentrations of these two species, is illustrated in fig. S24. Apparently, at lower temperature, the equilibrium between $HCo(CO)_4$ and $Co_2(CO)_8$ was shifted to the direction of the dimer.

We did not observe any decomposition of $HCo(CO)_4$ in the course of preformation under the applied conditions. In a stability test monitored with in situ IR spectroscopy at a cobalt loading of 5 mM [285 parts per million (ppm)] and 50-bar syngas (25 bar of which was CO partial pressure), the concentration of this active species did not decrease in the first 5 hours, 72% of which remained after 115 hours at 140°C (Fig. 2A and fig. S30). Even under 20-bar syngas we recorded the concentration of this active species formed from $Co_2(CO)_8$ in *n*-dodecane during a time span of 20 hours at the same temperature (Fig. 2E). No decline in concentrations of $HCo(CO)_4$ and $Co_2(CO)_8$ was observed. The two compounds coexisted in the solution in a steady-state equilibrium after the preformation was accomplished, in which the active catalyst, $HCo(CO)_4$, dominated in about 80.5% of total cobalt concentration (table S13). The common decomposition product $Co_4(CO)_{12}$ due to insufficient CO/H_2 pressure was not observable. This result accounts for the reliability of the hydroformylation catalysis at 30- to 50-bar syngas and 140°C.

To ascertain the threshold of $HCo(CO)_4$ stability, we monitored catalyst preformation at 20-bar syngas and 160°C, starting from $Co_2(CO)_8$ (Fig. 2F). Substantial decomposition of both $HCo(CO)_4$ and $Co_2(CO)_8$ occurred under such conditions. Owing to insufficient CO partial pressure, $Co_2(CO)_8$ was, as expected, to a small extent further converted to the $Co_4(CO)_{12}$ cluster. The highest concentration of the latter was 0.18 mM. The overall loss of soluble cobalt species, including $Co_2(CO)_8$, $HCo(CO)_4$, and $Co_4(CO)_{12}$, was about 59% after 21 hours, which means that starting from a solution of cobalt concentration of 5 mM, only 2 mM of cobalt was left in the homogeneous phase after the experiment. Thus, the unmodified cobalt catalyst, while remaining stable at 20-bar syngas and 140°C, cannot preserve its structure when the temperature is raised to 160°C. A higher CO partial pressure is required to sustain catalytic activity. As is shown in figs. S29 and S31, when the syngas pressure was increased to 50 bar, a stable catalyst system resulted, with a concentration of $HCo(CO)_4$ of 4.9 mM (table S11). We have in this way preliminarily discerned the threshold of the reaction conditions that can stabilize the active species in the phosphorus-free cobalt-based catalyst system. This result accounts for the findings of Bourne and co-workers (*24*), in which propylene was converted to aldehyde in the presence of 7.3 mM $Co_2(CO)_8$ under 35-bar syngas (H_2:CO = 2.5:1) at 110°, 130°, and 150°C, respectively. Unmodified cobalt carbonyl is not only active under such mild conditions for a short period but can also retain a long-term stability. These observations underpin the fact that the stability and feasibility of this classic catalyst has been considerably underestimated.

On the basis of these stability tests, we explored hydroformylation of three sterically hindered internal olefins under 30 bar of syngas at 140°C with the unmodified cobalt carbonyl complex. The results (Table 2) show that $HCo(CO)_4$ remained stable and active for 20 hours of reaction time, affording 100% linear selectivity for the branched 2,4,4-trimethyl-2-pentene and 2,3-dimethyl-2-butene substrates (entries 1 and 3, Table 2).

It is well accepted that electron-rich phosphine ligands enhance cobalt back-bonding to carbonyl, inhibiting CO dissociation and therefore the rate of hydroformylation (*2*, *13*). We compared the yields of aldehyde produced from hydroformylation of 1-octene when varied quantities of DPPE (1,2-bis(diphenylphosphino) ethane) were added along with the $Co(acac)_2$

catalyst precursor. An immediate decline of the activity was not observed. By contrast, we found a slight promoting effect of bisphosphines on the hydroformylation activity (fig. S34). This phenomenon was also confirmed by our further catalytic tests with other precursors (table S15). Regardless of whether the presynthesized catalyst **Co3**, with a phosphorus-cobalt ratio of 0.23 (for preparation, see the supplementary materials), or its in situ–generated equivalent was used, higher yields of aldehyde were observed when compared with the unmodified catalyst (entries 1 to 4, table S15). This effect was reversed when the P/Co ratio was increased to 0.9 (fig. S34). We attribute this promotion of activity to an operative effect of two catalytic species, $HCo(CO)_4$ and $HCo(CO)_2$(DPPE), which coexisted in solution at a low phosphorus loading. An increase in the amount of phosphine resulted in a dominating population of P-modified cobalt carbonyl, hampering catalytic activity. Frediani and coworkers reported similar observation from hydroformylation of 1-hexene with $Co_2(CO)_6L_2$ (L, alkyl phosphines) in the presence of free ligand, under 5 bar of CO and 40 bar of H_2 at 150°C (*20*).

The insights here into the stability of unmodified cobalt carbonyls motivate further study for application of this catalyst.

REFERENCES AND NOTES

1. O. Roelen, *Verfahren zur Herstellung von sauerstoffhaltigen Verbindungen*, German Patent DE849 548 (1938/1952).
2. A. Börner, R. Franke, *Hydroformylation. Fundamentals, Processes, and Application in Organic Synthesis* (Wiley-VCH, 2016), vols. 1 and 2.
3. R. Franke, D. Selent, A. Börner, *Chem. Rev.* **112**, 5675–5732 (2012).
4. A. C. Brezny, C. R. Landis, *Acc. Chem. Res.* **51**, 2344–2354 (2018).
5. A. Behr, P. Neubert, *Applied Homogeneous Catalysis* (Wiley-VCH, 2012).
6. P. W. N. M. van Leeuwen, C. Claver, *Rhodium Catalyzed Hydroformylation* (Kluwer Academic Publishers, 2002).
7. S. Li *et al*., *Org. Lett.* **22**, 1108–1112 (2020).
8. C. You *et al*., *J. Am. Chem. Soc.* **140**, 4977–4981 (2018).
9. X. Jia, Z. Wang, C. Xia, K. Ding, *Chemistry* **18**, 15288–15295 (2012).
10. M. Janssen, J. Wilting, C. Müller, D. Vogt, *Angew. Chem. Int. Ed.* **49**, 7738–7741 (2010).
11. G. Achonduh, Q. Yang, H. Alper, *Tetrahedron* **71**, 1241–1246 (2015).
12. B. Cornils, in *New Syntheses with Carbon Monoxide*, J. Falbe, Ed. (Springer, 1980), pp. 1–181.
13. F. Hebrard, P. Kalck, *Chem. Rev.* **109**, 4272–4282 (2009).
14. F. Ungváry, L. Markó, *J. Organomet. Chem.* **20**, 205–209 (1969).
15. F. Ungváry, L. Markó, *J. Organomet. Chem.* **193**, 383–387 (1980).
16. R. W. Wegman, T. L. Brown, *J. Am. Chem. Soc.* **102**, 2494–2495 (1980).
17. R. J. Klingler, M. J. Chen, J. W. Rathke, K. W. Kramarz, *Organometallics* **26**, 352–357 (2007).
18. R. J. Klingler, J. W. Rathke, *J. Am. Chem. Soc.* **116**, 4772–4785 (1994).
19. L. H. Slaugh, R. D. Mullineaux, *J. Organomet. Chem.* **13**, 469–477 (1968).
20. L. Rosi, A. Bini, P. Frediani, M. Bianchi, A. Salvini, *J. Mol. Catal. Chem.* **112**, 367–383 (1996).
21. D. M. Hood *et al*., *Science* **367**, 542–548 (2020).
22. *Verfahren zur Herstellung von Aldehyden durch die Oxosynthese*, German Patent DE 1 218 428 (1956).
23. T. K. Banerjee, C. R. Saha, D. Sen, *J. Chem. Technol. Biotechnol.* **31**, 737–744 (1981).
24. R. V. Gholap, O. M. Kut, J. R. Bourne, *Ind. Eng. Chem. Res.* **31**, 1597–1601 (1992).
25. M. Haumann, R. Meijboom, J. R. Moss, A. Roodt, *Dalton Trans.* **11**, 1679–1686 (2004).
26. R. Meijboom, M. Haumann, A. Roodt, L. Damoense, *Helv. Chim. Acta* **88**, 676–693 (2005).
27. L. Alvila, T. A. Pakkanen, T. T. Pakkanen, O. Krause, *J. Mol. Catal.* **73**, 325–334 (1992).
28. J. E. Dengler, A. Doroodian, B. Rieger, *J. Organomet. Chem.* **696**, 3831–3835 (2011).
29. Y. L. Borole, R. V. Chaudhari, *Ind. Eng. Chem. Res.* **44**, 9601–9608 (2005).
30. J. L. Arias *et al*., *Trans. Met. Chem. (Weinh.)* **38**, 787–792 (2013).
31. M. Vilches-Herrera, L. Domke, A. Börner, *ACS Catal.* **4**, 1706–1724 (2014).
32. R. F. Heck, D. S. Breslow, *J. Am. Chem. Soc.* **82**, 4438–4439 (1960).
33. R. F. Heck, D. S. Breslow, *J. Am. Chem. Soc.* **83**, 4023–4027 (1961).
34. C.-F. Huo, Y.-W. Li, M. Beller, H. Jiao, *Organometallics* **22**, 4665–4677 (2003).
35. E. Veleckis, D. S. Hacker, *J. Chem. Eng. Data* **29**, 36–39 (1984).
36. C. Kubis *et al*., *Chemistry* **20**, 11921–11931 (2014).
37. A. van Rooy *et al*., *Organometallics* **15**, 835–847 (1996).
38. K. Neymeyr, M. Sawall, D. Hess, *J. Chemometr.* **24**, 67–74 (2010).
39. M. Sawall *et al*., *J. Iran. Chem. Soc.* **13**, 191–205 (2016).

ACKNOWLEDGMENTS

We are indebted to A. Börner (LIKAT Rostock) for critical discussions and comments concerning this research and the manuscript. **Funding:** The authors gratefully acknowledge the financial support from Evonik Operations GmbH. **Author contributions:** Conceptualization: B.Z., C.K., R.F. Catalyst preparation, catalysis investigation: B.Z. Kinetic studies, in situ IR studies: B.Z., C.K. Funding acquisition, project initiation: R.F. Writing: B.Z., C.K., R.F. **Competing interests:** B.Z., C.K., R.F. are inventors on European patent application EP21177380. B.Z. and R.F. are inventors on European patent application EP21177383. R.F. is employee of Evonik Operations GmbH. **Data and materials availability:** All data are available in the main text or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm4465
Materials and Methods
Figs. S1 to S35
Tables S1 to S15
Eq. S1

Submitted 17 September 2021; resubmitted 28 November 2021
Accepted 9 August 2022
10.1126/science.abm4465

SOLAR CELLS

Transporting holes stably under iodide invasion in efficient perovskite solar cells

Tao Wang[1,2,3], Yao Zhang[1], Weiyu Kong[1,2,3], Liang Qiao[1,2,3], Bingguo Peng[1,2,3], Zhichao Shen[1], Qifeng Han[1], Han Chen[1], Zhiliang Yuan[4], Rongkun Zheng[5], Xudong Yang[1,2,3]*

Highly efficient halide perovskite solar cells generally rely on lithium-doped organic hole transporting layers that are thermally and chemically unstable, in part because of migration of iodide anions from the perovskite layer. We report a solution strategy to stabilize the hole transport in organic layers by ionic coupling positive polymer radicals and molecular anions through an ion-exchange process. The target layer exhibited a hole conductivity that was 80 times higher than that of the conventional lithium-doped layer. Moreover, after extreme iodide invasion caused by light-soaking at 85°C for 200 hours, the target layer maintained high hole conductivity and well-matched band alignment. This ion-exchange strategy enabled fabrication of perovskite solar cells with a certified power conversion efficiency of 23.9% that maintained 92% under standard illumination at 85°C after 1000 hours.

Metal halide perovskite solar cells (PSCs) with high power conversion efficiencies (PCEs) generally incorporate lithium salt–doped organic hole transporting layers (HTLs) (*1*–*4*) that can cause severe device instability issues (*5*, *6*). For example, the hygroscopic lithium salts can cause moisture invasion, migration of small Li^+ ions into perovskite, and irreversible device degradation (*7*–*10*). Inorganic HTLs and dopant-free HTLs offer great enhancements in stability but have failed to compete with the doped HTLs for the most efficient PSCs.

An additional stability issue arises from the intrinsic soft nature of the perovskites that allows volatile iodide components to enter charge-transporting layers (*11*, *12*). Iodide can interact with the positive radicals in HTLs and induce rapid degradation in both hole conductivity and interfacial band alignment, especially for the most efficient PSCs that rely on doped organic HTLs (*3*, *4*, *6*, *13*–*16*). Although iodide blocking barriers and interface engineering are promising strategies to hinder the iodide invasion (*11*, *12*), these approaches have not been reported to block both Li^+ migration and iodide invasion completely. Thus, it is highly desirable to find a Li-free doping

[1]State Key Laboratory of Metal Matrix Composites, Shanghai Jiao Tong University, Shanghai 200240, China. [2]Center of Hydrogen Science, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. [3]Zhangjiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai 201210, China. [4]Beijing Academy of Quantum Information Sciences, Beijing 100193, China. [5]School of Physics, Sydney Nano Institute, The University of Sydney, Sydney, NSW 2006, Australia.
***Corresponding author. Email: yang.xudong@sjtu.edu.cn**

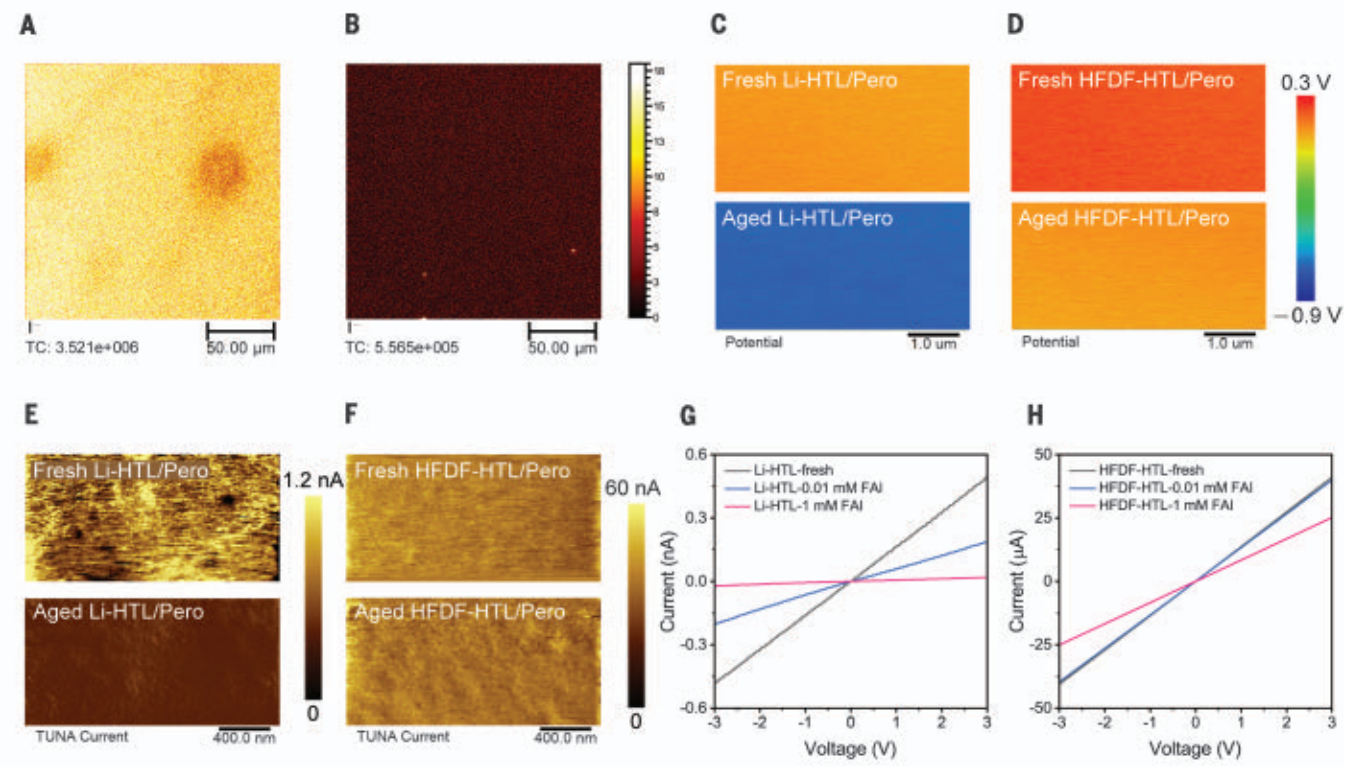


Fig. 1. Degradation of Li-HTL and HFDF-HTL under iodide invasion. (**A** and **B**) 2D TOF-SIMS elemental mapping of I^- signals in Li-HTL (A) and HFDF-HTL (B) after aging. TC, total counts. (**C** and **D**) *xy*-plane surface potential mapping images of Li-HTL/perovskite (C) and HFDF-HTL/perovskite (D) before (top) and after (bottom) aging. Pero, perovskite. (**E** and **F**) Variations in the surface conductivity of Li-HTL/perovskite (E) and HFDF-HTL/perovskite (F) before (top) and after (bottom) aging. The devices were all aged at the maximum power point under light-soaking (AM1.5G, 100 mW cm^{-2}) for 200 hours at 85°C. TUNA, tunneling AFM. (**G** and **H**) The current-voltage curves of Li-HTL (G) and HFDF-HTL (H) before and after spin-coating with 0.01 and 1 mmol $liter^{-1}$ FAI solution in acetonitrile at a speed of 5000 rotations per minute (rpm).

strategy that can stabilize the hole transport under mass iodide entry.

We report a solution strategy to stabilize the hole transport in Li-free organic layers by coupling positive polymer radicals and molecular anions through an ion-exchange process (see methods for details). The molecular anion of 1,1,2,2,3,3-hexafluoropropane-1,3-disulfonimide ($HFDF^-$) (fig. S1) was exchanged into poly(triarylamine) (PTAA) films and termed HFDF-HTL (fig. S2). $HFDF^-$ has not been used as a noncoordinating ion in a p-type organic semiconductor, as far as we know. We also chose bis(trifluoromethane)sulfonimide lithium salt/4-*tert*-butylpyridine (LiTFSI/tBP)–doped poly(triarylamine) (Li-HTL) as a reference HTL that is generally used for highly efficient PSCs and offers better stability than LiTFSI/tBP-doped 2,2′,7,7′-tetrakis(*N*,*N*-di-p-methoxyphenylamine)-9,9′-spirobifluorene (Li-Spiro-OMeTAD).

The target Li-free–doped HTL, HFDF-HTL, increased hole conductivity to a value 80 times greater than that of the conventional lithium salt–doped HTLs, had a favorable energy-level alignment at the HTL/perovskite heterojunction, and achieved a high conductivity on fresh HFDF-HTL after a few minutes versus the conventional Li-salt doping method that generally takes several hours or days. Under the mass invasion of iodide, hole transport was stabilized through the ionic coupling of positive polymer radicals with introduced molecular anions. Furthermore, hole transport through organic layers withstood thermal stress at temperatures up to 120°C. These results helped us obtain efficient and stable PSCs that achieved a certified PCE of 23.9% and maintained 92% after operation at the maximum power point under AM1.5G solar light (100 mW cm^{-2}) at 85°C after 1000 hours.

We studied the electronic properties of the aged HTLs in PSCs, with a structure of fluorine-doped tin dioxide (FTO)/SnO_x/SnO_2/formamidinium–based perovskite/HTL/Au, that suffered severe iodide invasion after light-soaking at 85°C for 200 hours. The iodide invasion into the aged HTLs, as studied by time-of-flight secondary ion mass spectrometry (TOF-SIMS) chemical imaging (Fig. 1, A and B), showed high concentrations of I^- ions distributed over the aged HTLs. The compositional depth profiles of the aged PSCs further revealed that the iodide invasion started from the perovskite/HTL interface (fig. S3). The surface potential measured with Kelvin probe force microscopy (KPFM) (Fig. 1, C and D) of the I^--rich Li-HTL decreased from an initial 92 mV to a final −678 mV. By contrast, the surface potential of the I^--rich HFDF-HTL has a much smaller decrease from 202 to 102 mV, and the decreased value was nearly one-eighth that of the I^--rich Li-HTL, exhibiting astonishing durability of HFDF-HTL under iodide invasion during the heat and light aging test. We also characterized the electronic structure of I^--rich HTLs with ultraviolet photoelectron spectroscopy (UPS). The Fermi level of the I^--rich Li-HTL in aged devices sharply changed from −4.96 to −4.02 eV (fig. S4), which reflected n-doping caused by iodide migration during the entire aging process. However, only a slight change from −5.12 to −4.97 eV in the Fermi level of I^--rich HFDF-HTL was found after the aging test.

The variation of electronic conductivity of I^--rich HTLs of aged PSCs was measured with conductive atomic force microscopy (AFM) (Fig. 1, E and F). The average electric current of I^--rich Li-HTL quickly decreased from the initial 0.76 to 0.43 nA and maintained only

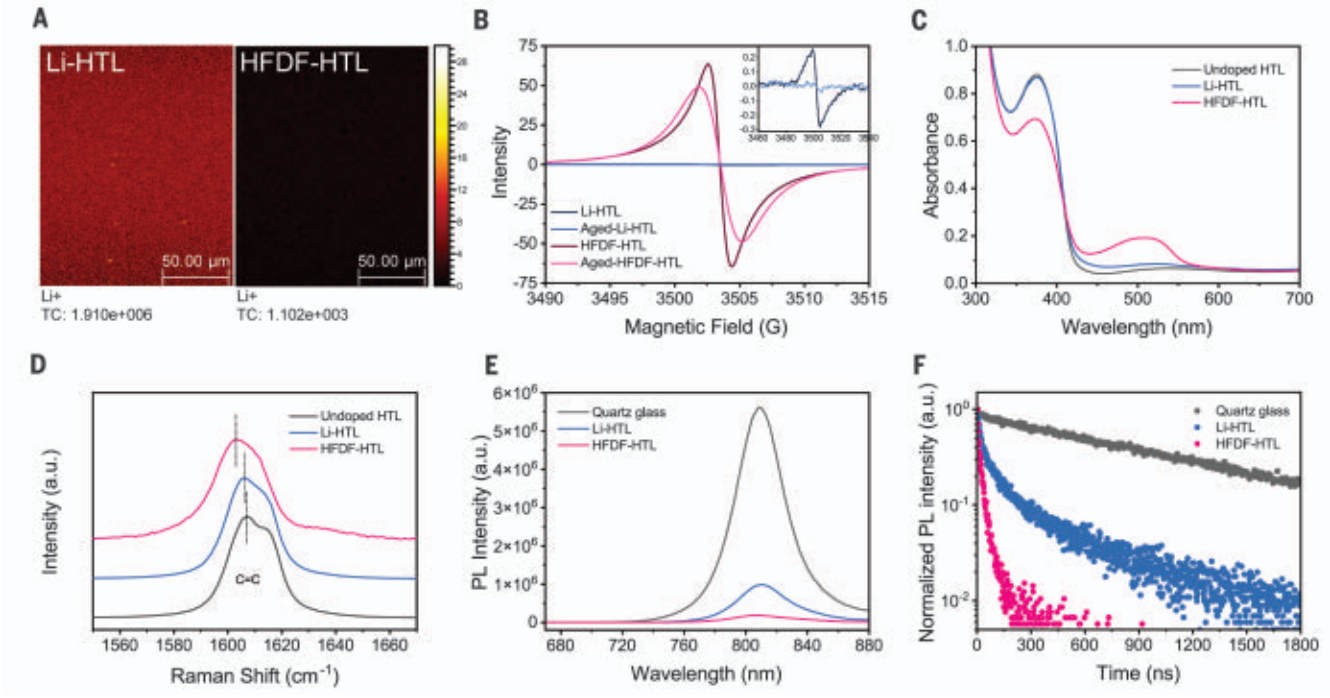


Fig. 2. Li^+ and radical content and charge transport properties of Li-HTL and HFDF-HTL. (**A**) Mapping of TOF-SIMS signals of Li^+ cations in Li-HTL and HFDF-HTL. (**B**) ESR spectra of Li-HTL and HFDF-HTL before and after treatment with 0.01 mmol liter^{-1} FAI solution. The inset shows a zoomed-in view of the low-intensity region. (**C** and **D**) The optical absorbance (C) and Raman spectra within the range of 1550 to 1670 cm^{-1} (D) of the undoped HTL, Li-HTL, and HFDF-HTL. (**E** and **F**) Steady PL spectra (E) and TRPL decay profiles (F) of perovskite films with and without the different doped HTLs. a.u., arbitrary units.

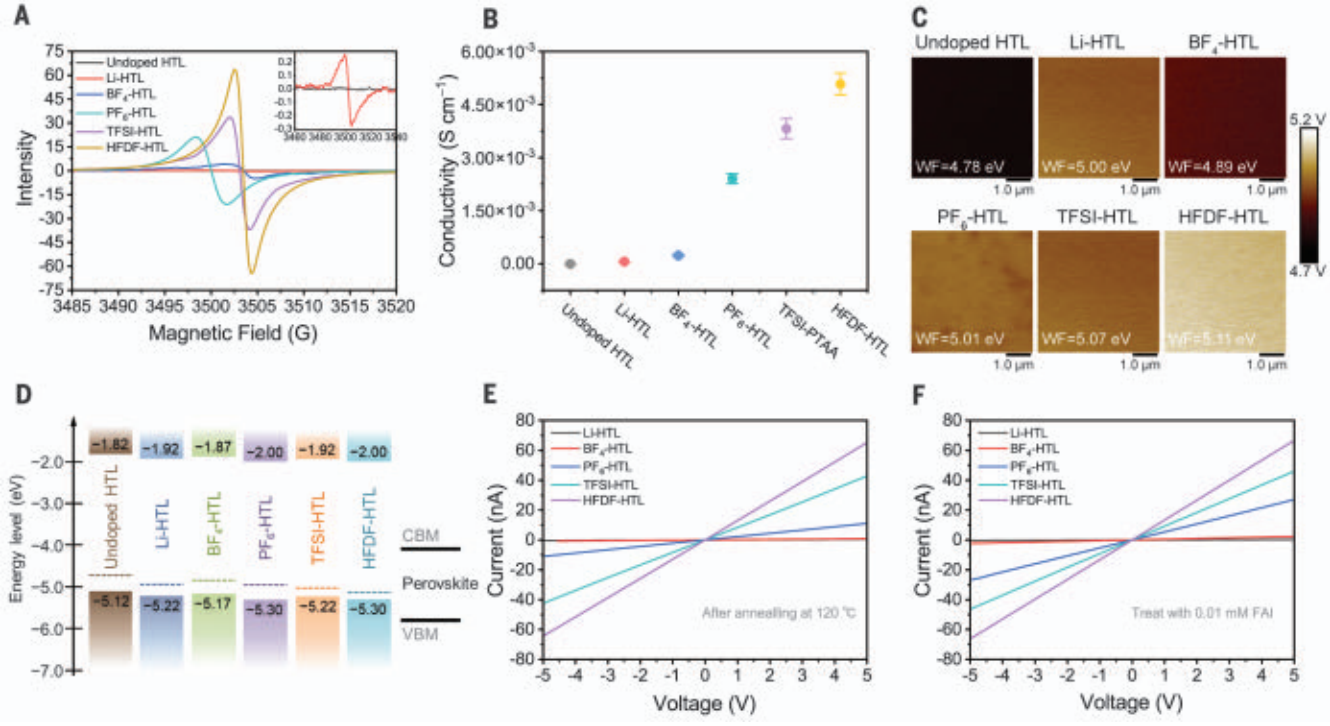


Fig. 3. Hole-related properties of the pristine and different doped HTLs. (**A**) ESR spectra of the undoped HTL, Li-HTL, and ion-exchanged HTLs with different lithium salts. The inset shows a zoomed-in view of the low-intensity region. (**B**) Variations in the electrical conductivity of the different HTLs. The error bars in the conductivity represent the standard deviation. (**C**) The *xy*-plane work function mapping images of the different HTLs. The surface work function (WF) was calibrated by Au foil. (**D**) Energy-level diagrams for the different HTLs and the perovskite layer based on the parameters derived from UPS spectra. CBM, conduction band minimum; VBM, valence band maximum. (**E** and **F**) The current-voltage curves of the different HTLs after annealing at 120°C for 1 day (E) and after spin-coating with 0.01 mmol liter^{-1} FAI acetonitrile solution at a speed of 5000 rpm (F).

57% of the initial conductivity. In comparison, I^--rich HFDF-HTL exhibited a stable performance with an initial value of 38.6 nA, which was ~50 times greater than that of the fresh Li-HTL, and the aged sample maintained 96% of the initial conductivity (36.9 nA). The conductivity of I^--rich HTLs was further investigated under the same iodide concentrations, which were aged through direct injection of formamidinium iodide (FAI), a common volatile perovskite component, into the layer by spin-coating. The conductivity of I^--rich Li-HTL decreased by 50% at the FAI concentration of 0.01 mmol liter^{-1} (Fig. 1G), whereas the conductivity of I^--rich HFDF-HTL remained at 97% under the same conditions (Fig. 1H).

We characterized the compositions of HTLs, especially the positive polymer radicals that can determine their conductivity. In Fig. 2A, Li-HTL was fully occupied by Li^+ signals, whereas only negligible Li^+ signals were observed in HFDF-HTL. We also confirmed the absence of Li^+ cations in HFDF-HTL by high-resolution x-ray photoelectron spectroscopy in fig. S5. From TOF-SIMS two-dimensional (2D) elemental mapping (fig. S6A) and compositional depth profiles (fig. S3B), we observed that $HFDF^-$ anions were distributed throughout the layer. TOF-SIMS signals of the anions of initiating dopant 2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane (F4TCNQ) were negligible in HFDF-HTL (fig. S6B), indicating that $F4TCNQ^-$ anions were efficiently exchanged with $HFDF^-$ and then removed during the spin-coating and washing process. We further verified the ion exchange for $HFDF^-$ according to the absence of the $F4TCNQ^-$ anions in the ion-exchanged HTL by Fourier transform infrared and Raman spectroscopy (figs. S7 and S8) (*17*). The positive polymer radicals were characterized by electron spin resonance (ESR) spectroscopy, which showed a stronger signal from HFDF-HTL than from Li-HTL (Fig. 2B), confirming the formation of a high concentration of polymer radicals in HFDF-HTL (*18*).

After being aged under iodide invasion, the positive polymer radicals in Li-HTL samples were almost entirely quenched, whereas HFDF-HTL maintained 95.5% of the original value deduced from the second integral of the ESR derivative curves (*19*). By observing the bleaching of the neutral absorption of HTL at 376 nm and swelling of the new absorption of polymer radicals at ~500 nm, an increase in polymer radicals in HFDF-HTL through ion exchange was further verified (Fig. 2C) (*20*). The Raman peak positions of the C=C stretching in conjugated polymers shifted to lower frequencies (Fig. 2D), indicating higher effective conjugation length and planarization of charged chains (*21*). The density functional theory (DFT) calculation (fig. S9) results supported the observation by Raman spectroscopy, in that the dihedral angles between bridging benzene rings of neutral and one and two positively charged triarylamine dimers were 143.7°, 172.4°, and 178.0°, respectively. The experimental observations of the electronic performance and compositions of HTLs show that the invaded I^- ions reduced the positive polymer radicals and caused degradation in hole conductivity for conventional Li-HTL. By contrast, the hole transport was well protected against the iodide invasion in HFDF-HTL.

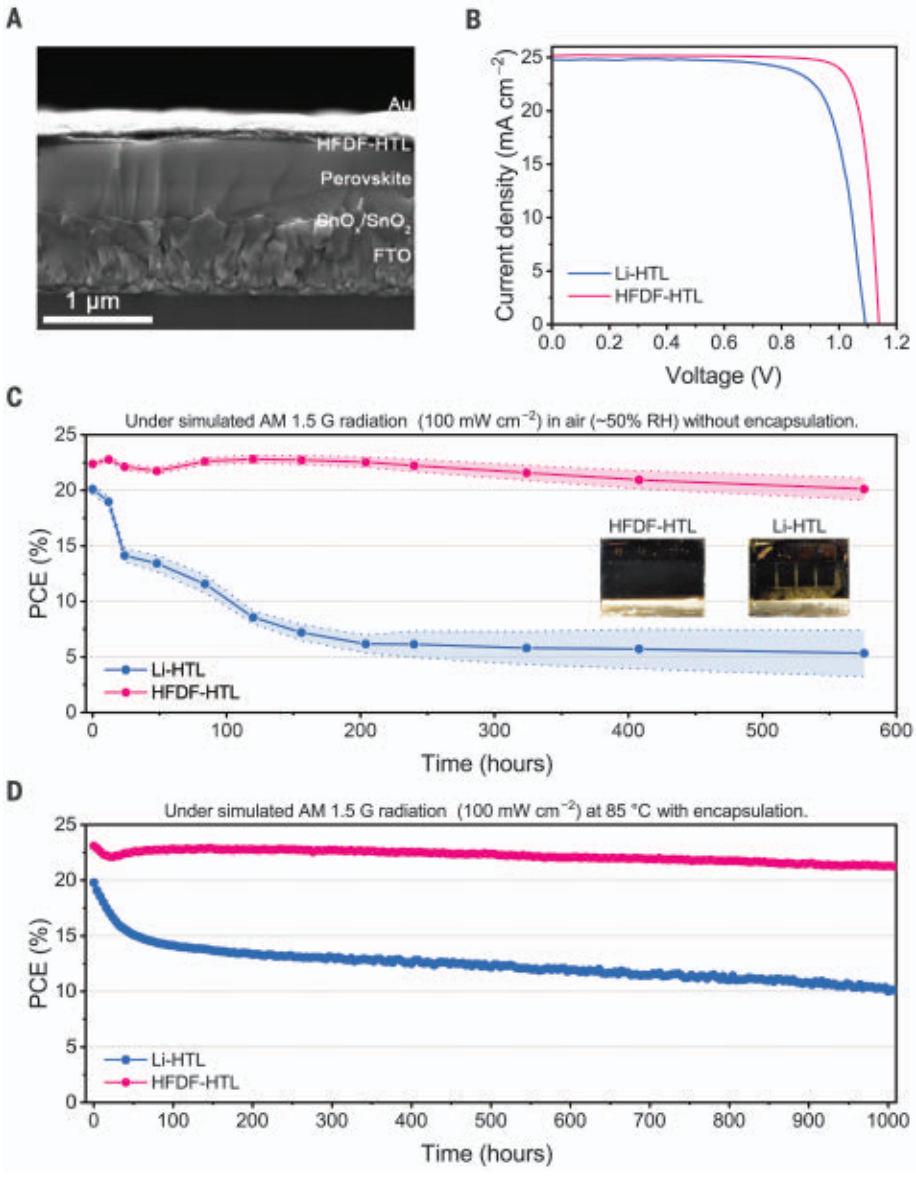


Fig. 4. Structure, photovoltaic performance, and long-term stability studies of the PSCs with Li-HTL or HFDF-HTL. (**A**) Cross-sectional scanning electron microscopy image of the PSCs. (**B**) Current density–voltage characteristics of the best-performing n-i-p devices with the different HTLs. (**C**) Long-term moisture stability of the Li-HTL– and HFDF-HTL–based devices under simulated AM1.5G radiation (100 mW cm^{-2}) in the air (~50% RH) without encapsulation. The shaded regions represent standard deviations from the average PCE obtained from eight aged cells. The insets show pictures of the devices after aging for 576 hours. (**D**) Long-term operational stability for the most-stable encapsulated devices among eight aged cells with the different HTLs under AM1.5G illumination (100 mW cm^{-2}) and heat stress at 85°C.

We found that HTLs doped with only F4TCNQ or lithium salts had no obvious enhancement of polymer radicals (fig. S10). For HFDF-HTL that was formed through an ion-exchange process with LiHFDF or KHFDF, both HTLs exhibited a similar electronic conductivity (fig. S11), indicating that the HFDF-HTLs are less sensitive to Li^+ or K^+ cations. In

addition, we found that HFDF-HTLs that were formed through the ion-exchange method had a higher radical doping effect than those fabricated by immersion from the observations of higher characteristic absorption at ~500 nm and much higher conductivity (fig. S12). The spin-coating drop by drop could provide fresh dopant solution continuously for ion-exchange processing and lead to a high-speed washing effect that cleared away some by-products. We also assessed the morphological change of HTL after doping by AFM (fig. S13). Pinholes were observed in Li-HTL, whereas the surface of HFDF-HTL was dense and uniform over the perovskite layer (fig. S14).

The effect of different doping strategies on the band alignment of HTLs was further evaluated (fig. S15). The Fermi level deduced from the value of high-binding-energy cut-off downshifted from −4.72 eV for undoped HTL to −4.96 eV for Li-HTL and −5.12 eV for HFDF-HTL, suggesting that the ion-exchange strategy enabled more-efficient p-type doping. The valence band maxima of Li-HTL and HFDF-HTL were estimated to be −5.22 and −5.30 eV, respectively, gradually approaching the value of perovskite film (fig. S16). The band alignment of HFDF-HTL reduced the energy loss during charge transfer and increased the open-circuit voltage (V_{OC}) (*22*).

We performed steady-state photoluminescence (PL) spectroscopy and time-resolved photoluminescence (TRPL) spectroscopy to study the charge transfer kinetics between perovskite and the different HTLs (Fig. 2, E and F). Table S1 summarizes the lifetime of the decay curves fitted by bi-exponential components using the equation $Y = A_1\exp(-t/\tau_1) + A_2\exp(-t/\tau_2)$, where fast decay lifetime τ_1 and slow decay lifetime τ_2 are related to trap-assisted nonradiative and radiative recombination, respectively (*23*). For undoped HTL, the PL intensity and lifetime did not fall off noticeably, indicating a relatively inefficient carrier extraction process without doping (fig. S17) (*24*). The conventional doping strategy exhibited a limited PL quenching effect for Li-HTL, whereas a steady PL spectrum was almost completely quenched for HFDF-HTL (Fig. 2E). The fitted PL lifetimes (τ_1 and τ_2) of perovskite film deposited on bare quartz glass of 56 and 1160 ns, respectively, were shortened more for HFDF-HTL (25 and 230 ns) compared with Li-HTL (54 and 382 ns) (Fig. 2F), indicating that photogenerated charges were transferred more sufficiently for HFDF-HTL.

We applied the ionic coupling between anions and polymer radicals with the ion-exchange process to more molecular cations (fig. S1)—such as tetrafluoroborate (BF_4^-), hexafluorophosphate (PF_6^-), and bistrifluoromethanesulfonimidate ($TFSI^-$)—for the formation of HTLs with enhanced conductivity and stability. The ESR and optical absorbance spectra show that the formation of polymer radicals was greatly enhanced in the order of HFDF-HTL, TFSI-HTL, PF_6-HTL, and BF_4-HTL (Fig. 3A and fig. S18), indicating that these molecular anions preferentially couple with the positive polymer radicals. The ionic coupling between anions and radicals was further evaluated with DFT calculations based on the energetic gain (figs. S19 and S20 and tables S2 to S4), which showed that $HFDF^-$ anions were more energetically favorable to couple with polymer radicals than $TFSI^-$ followed by PF_6^- and BF_4^-, hence higher ion-exchange efficiency and polymer radical concentration.

We measured the conductivity and hole mobility of the different HTLs (figs. S21 and S22). HFDF-HTL exhibited the highest electrical conductivity (Fig. 3B) followed by TFSI-HTL, PF_6-HTL, BF_4-HTL, and then Li-HTL. The conductivity of HFDF-HTL, which reached a value of 5.08×10^{-3} S cm^{-1}, was ~80 times higher than that of Li-HTL (table S5). HFDF-HTL also exhibited the highest hole mobility (fig. S23), which increased by nearly six times compared with that of Li-HTL (table S6). The exact distribution of charge carriers on the ion-exchanged HTLs and how their interaction with counterions affects electronic mobility pend further study. We further verified the work function in the different HTLs by KPFM (Fig. 3C). The energy band alignment of the different HTLs and the perovskite film (Fig. 3D) shows that HFDF-HTL had an optimal energy-level alignment at the HTL/perovskite interface.

The thermal stability of the HTLs was evaluated by measuring the conductivity before and after annealing at 80° or 120°C for 24 hours (figs. S24 and S25). The resulting current-voltage curves of the various HTLs annealed at 120°C showed that the thermally aged HFDF-HTL maintained 95% of the initial conductivity (Fig. 3E). The conductivity of Li-HTL films was reduced to 55% of initial values after heating at 120°C, possibly because of the undesired volatilization of tBP molecules and spontaneous aggregation of LiTFSI (*25*). By contrast, ion exchange improved the thermal stability in TFSI-HTL, maintaining 83% of the initial conductivity. The thermal stability of the conductivity was also consistent with the results of the DFT calculation on the cohesive energy (figs. S9B, S20, and S26 and tables S7 to S9), which indicated that the ionic coupling between anions and polymer radicals required a large activation energy for the ion decoupling under thermal stress (*26*, *27*).

The stability of HTLs under iodide invasion shows that the aged HFDF-HTL maintained the highest conductivity (Fig. 3F and figs. S26 and S27), followed by the aged TFSI-HTL, PF_6-HTL, BF_4-HTL, and then Li-HTL. Previous research reported that the diffusion of iodide ions from perovskite can reduce the oxidized HTLs (*28*, *29*). Both the thermal and chemical stability indicated that the ionic coupling between $HFDF^-$ anions and positive polymer radicals is energetically favorable and stabilizes the hole conductivity in HTLs against the reduction effect from the invaded iodide. Moreover, HFDF-HTL also enabled enhanced moisture stability of perovskite through its hydrophobicity. The contact angle of water on Li-HTL was 62.2° but was 85.0° for HFDF-HTL (fig. S29), which lacked hygroscopic Li^+ cations. During an aging test in the air with a relative humidity (RH) of ~50% for 500 hours, the HFDF-HTL/perovskite film exhibited a stable crystal structure as determined from x-ray diffraction (fig. S30).

An n-i-p planar heterojunction PSC fabricated with HFDF-HTL (Fig. 4A) achieved a maximum PCE of 24.0%, with a V_{OC} of 1.14 V, a short-circuit current density (J_{SC}) of 25.2 mA cm^{-2}, and a fill factor (FF) of 83.7% (Fig. 4B and table S10). By contrast, the Li-HTL–based devices only obtained a maximum PCE of 20.5% with a V_{OC} of 1.09V, a J_{SC} of 24.8 mA cm^{-2}, and a FF of 76.0%. The statistical analysis of the device performance with the different HTLs demonstrated that the HFDF-HTL–based devices exhibited superior performance to other devices we tested for all photovoltaic metrics (fig. S31). The improvement in device performance could be attributed to the high hole conductivity and well-matched band alignment (*23*, *30*).

The best device further performed maximum power point tracking with a voltage of 1.00 V (fig. S32). The photocurrent stabilized at ~23.9 mA cm^{-2}, corresponding to a stabilized PCE of 23.9%. We sent one of our best target PSCs to an accredited independent photovoltaic test center for certification. The device obtained a certified PCE of 23.9% with a V_{OC} of 1.15 V, a J_{SC} of 25.1 mA cm^{-2}, and a FF of 83.0% (fig. S33) and compared favorably to other reported Li-free PSCs to date (table S11). Our results also established advantages of $HFDF^-$ compared with the conventional lithium and/or cobalt salts for hole transport (table S12). We further applied the ion-exchange strategy to other polymeric HTLs (fig. S34), such as poly(3-hexylthiophene-2,5-diyl) (P3HT) and poly-{2,1,3-benzothiadiazole-4,7-diyl[4,4-bis(2-ethylhexyl)-4H-cyclopenta[2,1-b:3,4-b′]dithiophene-2,6-diyl]} (PCPDTBT), and their devices also showed enhanced V_{OC} and FF.

We performed the device stability test to investigate the durability of photovoltaic cells under moisture or combined heat and light. The nonencapsulated HFDF-HTL–based device retained nearly 90% of its initial performance after the aging test for 576 hours under simulated AM1.5G radiation (100 mW cm^{-2}) in ambient air (~50% RH) (Fig. 4C). By contrast, the control devices with Li-HTL rapidly lost 60% of their initial PCE under the same aging

conditions after only 100 hours. The operational stability of PSCs under heat and light stress was conducted under full-spectrum sunlight at 85°C (Fig. 4D). For the control cell with Li-HTL, a 49% drop in PCE was observed after continuous operation for 1000 hours, whereas the devices with HFDF-HTL maintained 92% of their initial PCE over the same aging period and still performed with a PCE of 21.2%.

REFERENCES AND NOTES

1. M. Kim *et al*., *Science* **375**, 302–306 (2022).
2. H. Min *et al*., *Nature* **598**, 444–450 (2021).
3. J. Jeong *et al*., *Nature* **592**, 381–385 (2021).
4. J. J. Yoo *et al*., *Nature* **590**, 587–593 (2021).
5. N. Li, X. Niu, Q. Chen, H. Zhou, *Chem. Soc. Rev.* **49**, 8235–8286 (2020).
6. K. Rakstys, C. Igci, M. K. Nazeeruddin, *Chem. Sci.* **10**, 6748–6769 (2019).
7. I. Lee, J. H. Yun, H. J. Son, T.-S. Kim, *ACS Appl. Mater. Interfaces* **9**, 7029–7035 (2017).
8. W. Li *et al*., *J. Mater. Chem. A Mater. Energy Sustain.* **2**, 13587–13592 (2014).
9. Z. Li *et al*., *Energy Environ. Sci.* **10**, 1234–1242 (2017).
10. J. A. Dawson *et al*., *ACS Energy Lett.* **2**, 1818–1824 (2017).
11. Y. Wang *et al*., *Science* **365**, 687–691 (2019).
12. E. Bi *et al*., *Nat. Commun.* **8**, 15330 (2017).
13. J. Kong *et al*., *Nature* **594**, 51–56 (2021).
14. T. H. Schloemer, J. A. Christians, J. M. Luther, A. Sellinger, *Chem. Sci.* **10**, 1904–1935 (2019).
15. P.-K. Kung *et al*., *Adv. Mater. Interfaces* **5**, 1800882 (2018).
16. N. Sakai *et al*., *Nat. Mater.* **20**, 1248–1254 (2021).
17. Y. Yamashita *et al*., *Nature* **572**, 634–638 (2019).
18. M. Polat, M. Korkmaz, *Food Res. Int.* **37**, 293–300 (2004).
19. H. D. Foersterling, L. Stuk, *J. Phys. Chem.* **96**, 3067–3072 (1992).
20. T. Ye *et al*., *ACS Appl. Mater. Interfaces* **11**, 14004–14010 (2019).
21. A. E. Mansour *et al*., *Phys. Chem. Chem. Phys.* **24**, 3109–3118 (2022).
22. X. Yin, Z. Song, Z. Li, W. Tang, *Energy Environ. Sci.* **13**, 4057–4086 (2020).
23. M. Jeong *et al*., *Science* **369**, 1615–1620 (2020).
24. Q. Jiang *et al*., *Nat. Energy* **2**, 16177 (2016).
25. S. Wang *et al*., *Nano Lett.* **16**, 5594–5600 (2016).
26. W. Zhao, J. Ding, Y. Zou, C. A. Di, D. Zhu, *Chem. Soc. Rev.* **49**, 7210–7228 (2020).
27. J. Liu *et al*., *Adv. Mater.* **30**, 1704630 (2018).
28. J. Carrillo *et al*., *Adv. Energy Mater.* **6**, 1502246 (2016).
29. S. Kim *et al*., *Sci. Rep.* **7**, 1200 (2017).
30. N. J. Jeon *et al*., *Nat. Energy* **3**, 682–689 (2018).

ACKNOWLEDGMENTS

We acknowledge assistance in measurements from the Instrumental Analysis Center of Shanghai Jiao Tong University (China). We thank J. Ding for assistance with TOF-SIMS measurements, H. Li for assistance with AFM and KPFM measurements, Z. Bao and X. Li for assistance with scanning electron microscopy measurements, N. Zhang for assistance with UPS measurements, R. Wang for assistance with PL and TRPL measurements, and X. Ding for assistance with x-ray photoelectron spectroscopy measurements. We thank L. Han for helpful discussions. **Funding:** This work was funded by the National Key Research and Development Program of China (grant no. 2018YFB1500104), the National Natural Science Foundation of China (grant nos. 11574199 and 11911530142), the Shanghai Pilot Program for Basic Research-Shanghai Jiao Tong University, and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. **Author contributions:** X.Y. supervised the project. X.Y. and T.W. conceived the ideas of the project and designed the experiments. T.W., Y.Z., W.K., and L.Q. characterized the films and devices. T.W. and Y.Z. performed the DFT calculation. T.W. conducted the fabrication of PSCs and their certification. T.W., Y.Z., W.K., L.Q., B.P., and Z.S. optimized the photovoltaic performance of the devices together. Q.H., H.C., Z.Y., and R.Z. participated in the discussion. X.Y. and T.W. co-wrote the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq6235
Materials and Methods
Figs. S1 to S34
Tables S1 to S12
References (*31*–*53*)

Submitted 21 April 2022; accepted 29 July 2022
10.1126/science.abq6235

What's Your Next Career Move?

From networking to mentoring to evaluating your skills, find answers to your career questions on *Science* Careers

To view the complete collection, visit
ScienceCareers.org/booklets

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers online @sciencecareers.org

By **Nico Eisenhauer**

A walk in the office

As my head bobbed up and down on the Zoom screen, private messages began to pour in from my colleagues: "What are you doing there?" "Are you sitting on a ball?" "Are you walking?" I could feel their eyes on me—inquiring, interested, confused. Clearly, this was not a normal faculty meeting. I replied that I was walking on my new underdesk treadmill. I was trying to combat the stagnation that had enveloped my workdays during the pandemic. It was an unusual solution, I knew. But I hoped my colleagues would accept it, and maybe even find their own ways to get moving.

I didn't have to worry about workday inactivity earlier in my career. As a graduate student and postdoc, I went on frequent field excursions to collect soil samples in remote and rugged places. The fieldwork was physically challenging, a contrast to the lab tasks and desk work that were also part of my job. But when I started my position as a professor, fieldwork became a rarity.

During the pandemic the situation grew worse. My schedule filled up with one online meeting or seminar after the next. In the past, I had maintained at least some activity during my workday by walking or biking to meeting and seminar rooms spread across the city. But now that meetings were virtual, I "traveled" to them without budging from my desk.

"I was trying to combat the stagnation that had enveloped my workdays."

Although I had the luxury of a height-adjustable desk, which gave me the option to stand, I still felt stagnant and unbalanced, and I began to experience back pain. I tried recommendations for desk-bound workers, such as taking regular breaks and being physically active before and after work, but they weren't enough.

At home, I had occasionally tried to work on my wife's treadmill, which I had repurposed by putting a wooden board across the handles and my laptop on top. I thought that might be just what I needed in my campus office, so I bought a treadmill that's designed to go under a desk. I decided to dive in headfirst and spend my entire workday on the treadmill. It took me a couple of days to adjust, but I eventually got the hang of typing while walking. For me, the key to success was to keep going and to learn by doing.

More recently, I have even started to give talks and lectures while walking. This was challenging at first, but I'm now used to speaking on a treadmill. I've noticed that I gesture more freely, which results in a more lively presentation. I do stop the treadmill for the Q&A session, because my footsteps can make it difficult to hear the questions. I also find that it's easier to think on the spot when I'm not walking.

But otherwise, when I'm at my desk, I spend nearly all of my working hours in motion. In the first couple of weeks, I was a bit exhausted when I got home at night. But over time, I developed more stamina. My back pain went away and I came to enjoy the time at my desk—so much so that when I started to travel for conferences again, I missed my daily routine of walking.

And I needn't have worried about my colleagues' reactions. Except for one person who felt distracted by my movements on Zoom, most of the scientists I work with have sent me positive and reassuring feedback. Some have even bought treadmills of their own. That's helped me feel less self-conscious about my movement during meetings. I've also come to view my walking as a way to silently communicate a crucial message to my colleagues: There are ways to stay active despite remote working and online meetings.

Not everybody can afford a stand-up desk and treadmill for their office. But there are easy and inexpensive alternative options. For instance, you could go for walks during phone calls, or you could set up a stand-up desk using materials you already have. During my time as a postdoc, I did that by propping up my laptop on a diaper box on top of the dishwasher. Be creative: Your physical and mental health depends on it. ■

Nico Eisenhauer is a research group head at the German Centre for Integrative Biodiversity Research (iDiv) and a professor at Leipzig University. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

You're only an essay away!

Check out our Awards and Prizes that are currently seeking submissions.

Immunology
Submissions Opened: May 1, 2022
Submissions Close: Oct 1, 2022

Scientific Innovation & Entrepreneurship
Submissions Opened: Jun 1, 2022
Submissions Close: Nov 1, 2022

NOSTER & Science Microbiome Prize
Submissions to Open: Sep 15, 2022
Submissions Close: Jan 24, 2023

Science & PINS Prize for Neuromodulation
Submissions to Open: Nov 1, 2022
Submissions Close: Mar 15, 2023

Science Awards and Prizes seeks to find the next generation of early-career researchers on the cusp of changing the scientific landscape. Start your journey by finding the Award or Prize that fits you and find out how to apply today!

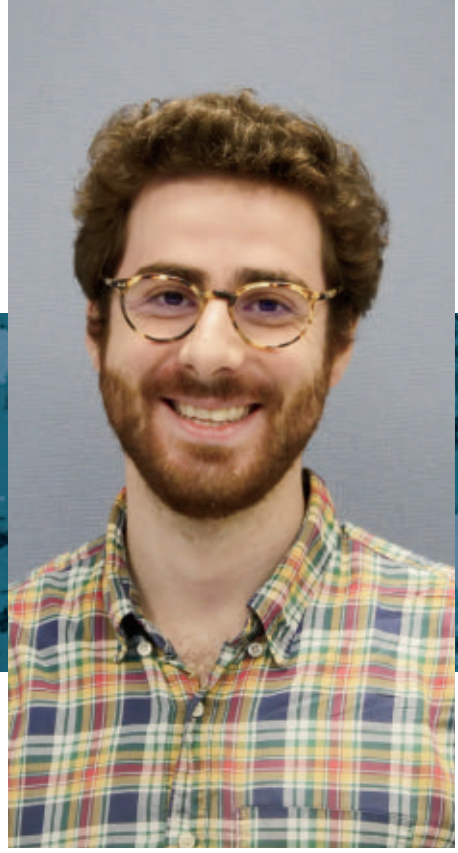
Science
AAAS

bit.ly/3qKxMN0

Scan the code to start your application process!

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY

REWARDING HIGH-RISK RESEARCH. SUPPORTING EARLY-CAREER SCIENTISTS. HELPING TO FIND CURES FASTER. **APPLY TODAY**



Now accepting applications for the Michelson Philanthropies & *Science* Prize for Immunology.

The Michelson Philanthropies and Science Prize for Immunology focuses on transformative research in human immunology, with trans-disease applications to accelerate vaccine and immunotherapeutic discovery. This international prize supports investigators 35 and younger, who apply their expertise to research that has a lasting impact on vaccine development and immunotherapy. It is open to researchers from a wide range of disciplines including computer science, artificial intelligence/machine learning, protein engineering, nanotechnology, genomics, parasitology and tropical medicine, neurodegenerative diseases, and gene editing.

Application deadline: Oct. 1, 2022.

For more information visit:
www.michelsonmedicalresearch.org

#MichelsonPrizes

"The Michelson Philanthropies & Science Prize for Immunology will greatly impact my future work. As I am just starting my scientific career, it will illuminate my work, spark interest and support me to continue my research in this field."

Paul Bastard, MD, PhD,
Laboratory of Human Genetics of Infectious Diseases, Imagine Institute (INSERM, University of Paris), Paris, France; and The Rockefeller University, New York.

Dr. Bastard received the inaugural Grand Prize for his essay: "Why do people die from COVID-19: Autoantibodies neutralizing type I interferons increase with age."

GRAND PRIZE:
$30,000

FINALIST PRIZE:
$10,000